

Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft

Band 71

Mit Daten Geschichten erzählen:
Das Potential von Datenjournalismus im
World Wide Web

Daniela Woytewicz

Oktober 2013

Fachhochschule Köln

Fakultät für Informations- und Kommunikationswissenschaften

Institut für Informationswissenschaft

Entstanden als Bachelorarbeit im Studiengang Online-Redakteur

Betreuer: Prof. Dr. Konrad Scherfer

Woytewicz, Daniela

Mit Daten Geschichten erzählen: Das Potential von Datenjournalismus im World Wide Web.

Köln: Fachhochschule Köln,

Fakultät für Informations- und Kommunikationswissenschaften,

Institut für Informationswissenschaft, 2013

(Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft; 71)

ISSN (elektronische Version) 1434-1115

Die Kölner Arbeitspapiere zur Bibliotheks- und Informationswissenschaft berichten über aktuelle Forschungsergebnisse des Instituts Informationswissenschaft der Fachhochschule Köln. Veröffentlicht werden sowohl Arbeiten der Dozentinnen und Dozenten als auch herausragende Arbeiten der Studierenden. Die Kontrolle der wissenschaftlichen Qualität der Veröffentlichungen liegt bei der Schriftleitung.

Jeder Band erscheint in elektronischer Version (über unsere Homepage:

<http://www.fbi.fh-koeln.de/institut/papers/arbeitspapiere.php>).

Fachhochschule Köln

Fakultät für Informations- und Kommunikationswissenschaften

Institut für Informationswissenschaft

Claudiusstr.1 D-50678 Köln

Tel.: 0221/8275-3376, Fax: 0221/3318583

E-Mail: schriftenreihe@fbi.fh-koeln.de

Schriftleitung: Prof. Dr. Hermann Rösch, Susanne Röltgen

© FH Köln 2013

ABSTRACT (DEUTSCH)

Im Zeitalter der fortschreitenden Digitalisierung und Technisierung von Informationen steigt der Datenberg im WWW kontinuierlich: Daten avancieren zum Rohstoff, mit dem sich viel Profit erzielen lässt. Gerade in Zeiten von Big Data und dem Kampf um die Informationshierarchie im WWW ist es für den Journalisten als Gatekeeper wichtig, aus der Fülle von Daten und Zahlen relevante Informationen herauszufiltern und sie der Öffentlichkeit verständlich zu präsentieren. Genau darin besteht die Herausforderung für das neue Feld des Online-Journalismus, den Datenjournalismus. Die vorliegende Arbeit geht dabei auf die Fragestellung ein, welches Potential der Datenjournalismus im WWW in sich birgt und wie er dabei helfen kann, eine komplexe, aus digitalen Daten bestehende Welt besser erklären zu können. Da der Datenjournalismus noch ein sehr junges journalistische Genre darstellt, wird er im Vorfeld der Arbeit zunächst definiert, historisch eingeordnet und reflektiert, um dadurch einen umfangreichen Einblick in das Thema zu ermöglichen. Anschließend werden die diversen Datentypen, die der Datenjournalismus als Quelle benutzt, näher erläutert und der Prozess Datenanalyse und -aufbereitung kurz angeschnitten. Anhand von Beispielen aus der Praxis wird auf die diversen Erzählformen und Methoden des Data-Driven-Journalismus eingegangen. Ziel der Arbeit ist es, die aufgeworfene Fragestellung mittels der Reflektion der vorher gewonnen Erkenntnisse zu beantworten und Zukunftsperspektiven für den Datenjournalismus aufzuzeigen.

ABSTRACT (ENGLISCH)

In the age of progressing digitalisation and mechanisation of information, the online data volume continuously grows. Data turns into a resource with huge profitability. Especially in times of Big Data and the constant struggle for information hierarchy in the WWW, it is important for the journalist, as the gatekeeper, to filter relevant information out of the bulk of data and to present them to the public in an understandable way. This is where the challenge for the new field of Online Journalism, Data Journalism, lies. This thesis will explore the potential Data Journalism has in the WWW and how it can help to better understand the complex world of digital data. Since Data Journalism is a rather new field, it will first be defined, historically classified and reflected upon in order to provide insight into the subject. Afterwards, the various types of data the Data Journalist uses as resources will be defined and the process of data analyses and data processing will be briefly looked into. Various examples will illustrate the methods of Data Driven Journalism. This thesis aims at answering the raised question by reflecting the findings and to present future perspectives for Data Journalism.

Schlagwörter: Datenjournalismus, Daten, Datenvisualisierungen, Big Data, Open Data, Crowdsourcing, interaktive Datenanwendungen, Online-Journalismus, Rohdaten, World Wide Web, Netz-Recherche

INHALTSVERZEICHNIS

1	EINLEITUNG.....	1
1.1	Problemhintergrund.....	1
1.2	Zielsetzung und Aufbau der Arbeit.....	2
1.3	Begriffserklärungen.....	2
2	DATA DRIVEN JOURNALISM	5
2.1	Die digitale Revolution	5
2.2	Information Overload – die Herausforderung	6
2.3	Definition Datenjournalismus	9
2.3.1	Charakteristika	13
2.3.2	Meilensteine des modernen Datenjournalismus.....	15
2.3.3	Journalistische Einordnung des Datenjournalismus.....	17
3	DATEN IM NETZ	19
3.1	Datengewinnung.....	19
3.1.1	Das Prinzip Open Data	20
3.1.2	Web scraping: Saugen aus dem Netz	22
3.1.3	Leaked Information	24
3.2	Datenanalyse und -aufbereitung.....	25
3.3	Erzählformen und Methoden.....	27
3.3.1	Data Storytelling	28
3.3.2	Echtzeitdaten	30
3.3.3	Datensätze	31
3.3.4	Crowdsourcing	32
3.3.5	Hyperlokal	37
4	DAS POTENTIAL DES DATENJOURNALISMUS	39
4.1	Exklusivität und Unabhängigkeit	39
4.2	Transparenz und Glaubwürdigkeit	41
5	FAZIT UND AUSBLICK.....	43
6	LITERATUR- UND QUELLENVERZEICHNIS	44

ABKÜRZUNGSVERZEICHNIS

API	Application Programming Interface
App	Applikation
CAR	Computer-Assisted Reporting
DDJ	Data Driven Journalism
DVD	Digital Versatile Disc
HTML	Hypertext Markup Language
PDF	Portable Document Format
SMS	Short Message Service
USA	United States of America
WWW	World Wide Web

ABBILDUNGSVERZEICHNIS

Abbildung 1: Faktoren zur Kennzeichnung einer datenjournalistischen Arbeit.....	14
Abbildung 2: Die Datenzwiebel	19
Abbildung 3: Visualisierung „Verräterisches Handy“ auf <i>zeit.de</i> auf Basis der Handydaten des Politikers Malte Spitz	29
Abbildung 4: Das Projekt „Zugmonitor“ auf <i>zeit.de</i>	31
Abbildung 5: Interaktive Datenbank über die Parteispenden von 2002 bis 2009 auf <i>zeit.de</i>	32
Abbildung 6: Das <i>Guardian</i> -Projekt “Investigate your MP’s Expenses” zum Spesenskandal in Großbritannien	34
Abbildung 7: Grafische Aufbereitung der im „GuttenPlag Wiki“ gesammelten Daten zu Karl-Theodor Freiherr zu Guttenbergs Dissertation „Verfassung und Verfassungsvertrag“	36
Abbildung 8: Vorreiterrolle im Bereich Hyperlokale-Werke: das Portal <i>Everyblock</i>	38

1 EINLEITUNG

1.1 Problemhintergrund

Die immer engere Vernetzung von Informationen über das Internet führt zu einem stetigen Wachstum der global verfügbaren Datenmasse, die kaum jemand mehr durchdringen kann. Gerade in diesen Zeiten des „Information-Overflows“ ist es für den Journalisten als Gatekeeper wichtig, aus der Fülle von Daten und Zahlen relevante Informationen herauszufiltern und sie der Öffentlichkeit verständlich zu präsentieren.

Genau darin besteht die Herausforderung für das neue Feld des Online-Journalismus, dem Datenjournalismus. Durch ihn kann die immer komplexer gewordene Gesellschaft anhand von Daten präziser, schneller und umfangreicher dargestellt werden. Dabei sind die Datensätze nicht nur Beiwerk oder Unterstützung zum Artikel: Beim Datenjournalismus ist der Datensatz Ausgangsbasis und Gegenstand der Berichterstattung zugleich; Daten dienen nicht mehr allein der Unterstützung einer These, sondern erst aus ihnen entsteht die eigentliche Fragestellung. Durch den Datenjournalismus werden die Daten aus dem World Wide Web systematisch ausgewertet, in interaktiven Grafiken visualisiert und die Originärquellen für die Leserschaft transparent gemacht, sowie Datensätze der Allgemeinheit zur eigenständigen Einsicht und Bearbeitung zur Verfügung gestellt.

Seit 2009 hat der Datenjournalismus eine steile Karriere hingelegt. So starteten die britische Tageszeitung *The Guardian* und die *New York Times* in diesem Jahr eigene Ressorts – den sogenannten Datastore und Data-Desk – die sich um das Aggregieren, Gewichten und Auswerten von Daten für redaktionelle Beiträge kümmern. Die beiden Medien gelten damit als Pioniere im Bereich des Data Driven Journalism. In den USA gehört der Datenjournalismus sogar zu den drei Informationssäulen: people, paper, data. Mit der Veröffentlichung der Afghanistan-Kriegstagebücher (Warlogs) im Jahr 2010 durch Wikileaks schaffte der Datenjournalismus endgültig seinen Durchbruch als eigenes Genre. Selten zuvor stand ein so großer Datensatz im Mittelpunkt des medialen Interesses und wurde einer breiten Masse verständlich zugänglich gemacht – und das vor allem durch die Möglichkeiten des Internets. So benutzte *The Guardian* die 90.000 militärischen Einsatzprotokolle aus Afghanistan von Julian Assanges Whistleblower-Portal nicht nur für herkömmliche und multimedial angereicherte Berichte, sondern erstellte auch eine für Nutzer in mehreren Ebenen recherchierbare Datenbank – ein Vorgehen, das international hoch gelobt wurde. Seit diesem Zeitpunkt schenkt der Journalismus, insbesondere der Online-Journalismus, dem Datenjournalismus immer mehr Beachtung. So stieg die Zahl datenjournalistischer Projekte im Internet in den letzten Jahren stetig an.

1.2 Zielsetzung und Aufbau der Arbeit

Diese gestiegene Aufmerksamkeit bietet ausreichend Anlass, die noch sehr junge Entwicklung des Datenjournalismus im World Wide Web zu untersuchen und sie systematisch abzubilden. Dabei geht die Arbeit der Frage nach, welches Potential der Datenjournalismus im World Wide Web für den (On-line-)Journalismus in sich birgt. Mit dem Titel der Arbeit „Mit Daten Geschichten erzählen“ ist nicht nur die Produktion eines fertigen Berichts gemeint, sondern schließt jegliche aus Daten gewonnenen Erkenntnisse, Ideen und Informationen ein, die im journalistischen Sinne bedeutungsvoll sind und deren Zusammenhänge zu interessanten, journalistischen Themen werden.

Zum besseren Verständnis der Thematik wird im Vorfeld ein theoretischer Rahmen geschaffen, der den Data Driven Journalism definiert, ihn historisch einordnet und reflektiert, sowie seine Grundlagen aufzeigt und dadurch einen umfangreichen Einblick in das Thema ermöglicht.

Im zweiten Teil der Arbeit wird gezeigt, zu welchem Zweck und mit welchen neuen Möglichkeiten das Potential der vorhandenen Datenmenge im Web durch den Datenjournalismus genutzt werden kann. Dafür wird als Erstes darauf eingegangen, woher und wie die Daten im Web gewonnen werden. In diesem Rahmen wird das Prinzip Open-Data näher erläutert und über die Technik des Webscrapens aufgeklärt. Auch wird hier auf Leaked-Information als Recherchequelle eingegangen. Anschließend wird der Prozess Datenanalyse und -aufbereitung kurz angeschnitten. Das Kapitel schließt mit einer näheren Betrachtung der diversen Erzählformen und Methoden des Data-Driven-Journalismus ab. Hierfür werden die datenjournalistischen Darstellungsformen in sechs verschiedene Typen unterteilt. Es werden Methoden, Funktion und Zweck der verschiedenen Bereiche Mittels Beispielen aus der Praxis erläutert.

Im Schlussteil der Arbeit werden die erworbenen Kenntnisse der vorherigen Kapitel reflektiert und daraus der Mehrwert des Datenjournalismus abgeleitet. Dieser Teil bildet einen fließenden Übergang zum Fazit, welcher die am Anfang aufgeworfene Frage, welches Potential der Datenjournalismus birgt, beantwortet und weiterführende Untersuchungen im Hinblick auf die Zukunftsperspektiven des Datenjournalismus aufzeigt.

1.3 Begriffserklärungen

Gatekeeper

Der Begriff wird in dieser Arbeit in Relation zur Nachrichtenforschung gesetzt. Unter Gatekeeper versteht man die Rolle von Journalisten, beziehungsweise Redaktionen zu entscheiden, welche Meinungen und Nachrichten veröffentlicht werden.

Whistleblower

Das Wort Whistleblower stammt aus dem Englischen “to blow the whistle” und meint wörtlich übersetzt „die Pfeife blasen“. Whistleblower beschreibt in dieser Arbeit Hinweisgeber oder Informanten, die Missstände wie zum Beispiel illegales Handeln an die Öffentlichkeit bringen.

API

Eine API erlaubt es anderen Softwareanwendungen, gezielt und automatisiert maschinenlesbare Daten abzufragen.

Internet

Obwohl das Internet fachlich gesehen die Infrastruktur in Form von Hardware und Protokollen für das World Wide Web ist, wird das Wort in dieser Arbeit als Synonym für das World Wide Web verwendet, da die beiden Begriffe in der hinzugezogenen Literatur im Allgemeinen nicht differenziert behandelt werden.

Web, Netz, WWW

Sind Synonyme für das World Wide Web.

Rezipient, Nutzer, Leser

Die Begriffe Leser, Nutzer und Rezipient werden in dieser Arbeit synonym verwendet. Sie sind Personen, die mediale Online-Angebote jeglicher Art konsumieren.

Informationen

Unter Informationen werden in dieser Arbeit digitale Informationen verstanden, die durch die Digitalisierung beliebiger Informationen entstehen und als Resultat Daten ergeben.

Daten

Meint in dieser Arbeit gruppierte Informationen. Sie sind Ansammlungen von Zeichen, die spezifische Informationen enthalten und elektronisch verarbeitet werden können.

Datensatz

Ist eine Ansammlung gleich strukturierter Daten mit mindestens einem Datenfeld (Zelle).

Rohdaten

Rohdaten sind unbearbeitete Informationen, die die Grundlage aller Datensätze bilden. „Die werden entweder unmittelbar – in Echtzeit oder Nachhinein – ‚roh‘ veröffentlicht (etwa von einer Messstelle per Schnittstelle/API). Wenn sie im Nachhinein bereinigt und bearbeitet bereitgestellt

werden, sind sie nicht mehr ‚roh‘ (beispielsweise ein Finanzhaushalt)“ (Matzat 2011e).

Scraper, Wrapper

Die Begriffe Scraper und Wrapper werden in dieser Arbeit synonym verwendet. Ein Scraper oder Wrapper ist ein Programm, das Daten aus einer Website oder einem Dokument ausliest (scraping). Andere geläufige Synonyme sind Webcrawler, Spider, Scraper, Webscraper, Suchagenten, Searchrobot, Wrapper, Searchbot, Bot, Webfetching, Web harvesting, Web data extraction oder webgrabber.

Outsourcing

Outsourcing ist eine Unternehmensstrategie, bei der das Unternehmen bestimmte Firmenbereiche, Geschäftsprozesse oder Aufgaben an spezielle Dienstleistungsunternehmen auslagert.

2 DATA DRIVEN JOURNALISM

2.1 Die digitale Revolution

Die Erfindung des Buchdrucks mit beweglichen Lettern durch Johannes Gensfleisch zu Gutenberg um das Jahr 1450 war zweifellos eine der größten Errungenschaften der Informationsgeschichte; Erstmals war man in der Lage Informationen und Wissen einem breiten Publikum zugänglich zu machen. Seitdem wächst auch die von Menschen produzierte Menge an Informationen (vgl. Diepold 1997). Wie Gutenbergs Erfindung revolutionierte auch die digitale Technik die Verbreitung von Informationen: Sie ist mit den Jahrzehnten immer schneller, der Zugang zu ihr immer einfacher und die Menge an zu bewältigenden Daten immer größer geworden (vgl. Rau 2002: 2).

Doch wie viele digitalisierte Informationen gibt es auf der Welt? Dieser Frage sind Martin Hilbert von der University of Southern California in Los Angeles und Priscila López von der katalonischen Universität Oberta mit Sitz in Santiago de Chile vier Jahre lang nachgegangen. Sie untersuchten, wie viele Informationen in den Jahren 1986 bis 2007 gespeichert, kommuniziert oder verarbeitet wurden. Die von den Wissenschaftlern im Jahr 2011 in der Zeitschrift *Science* veröffentlichte Studie belegt, dass die Menschheit 2002 erstmals mehr Informationen in digitaler Form gespeichert hat als in analoger (vgl. Wu 2011). Daher wird dieses Jahr auch als Beginn des digitalen Zeitalters angesehen und durch den Begriff digitale Revolution (auch elektronische Revolution) beschrieben. Während im Jahr 2000 noch 75 Prozent aller gespeicherten Informationen in analoger Form vorlagen (vor allem auf Videobändern), betrug der Anteil digitaler Daten 2007 bereits 97 Prozent. So wuchs die Datenkapazität in der Telekommunikation (hier verstanden als Telefon und World Wide Web) in nur sieben Jahren – von 2000 bis 2007 – um 28 Prozent pro Jahr und damit viermal schneller als die Weltwirtschaft (vgl. Bengsch 2011a). Damit konnten die Menschen im Jahr 2007 rund 295 Exabyte Informationen auf elektronischen Datenträgern lagern (vgl. Hilbert; Lopez 2011). Diese Menge entspricht der Kapazität einer Milliarde Computerfestplatten mit je 300 Gigabyte Potential. Seitdem wächst der digitale Informationsberg weiter, in den meisten Jahren exponentiell.

Laut einer aktuellen Studie *Cisco Visual Networking Index: Forecast and Methodology, 2010-2015* (Cisco 2011) des US-amerikanischen Telekommunikationsunternehmens Cisco wird sich der weltweite Datenverkehr über das Internet bis zum Jahr 2015 im Vergleich zu dem Volumen aus 2010 vervierfachen: „Insgesamt wird im Jahr 2015 fast ein Zettabyte an Daten über IP-Netze wie das Internet übertragen. Das ist eine Eins mit 21 Nullen und entspricht den Inhalten von sechs Milliarden DVDs pro Minute“ (Köppli 2011). Auch im Hinblick auf den einzelnen Privatanwender steigen die Zahlen: Hat ein einzelner Internetnutzer im Jahr 2010 durchschnittlich 7,3 GB pro Monat übertragen, werden es laut Studie 2015 bereits 24,8 GB

pro Monat sein. Dabei nimmt vor allem der Datenverkehr bei Privatanwendern mit einer konstanten jährlichen Rate von 34 Prozent im Gegensatz zu dem der beruflichen Nutzer schneller zu, die nur 19 Prozent jährlich übertragen.

In dem darauf folgenden *White Paper*, dem *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016* (Cisco 2012) aus dem Jahr 2012 wird zusätzlich zu den Ergebnissen aus der Vorstudie erläutert, dass Internetnutzer Informationen vermehrt unterwegs, also über mobiles Internet, konsumieren werden. So werde, laut Prognose, der globale mobile Datenverkehr von 2011 bis 2016 um das Achtzehnfache wachsen und im Jahr 2016 über 130 Exabyte mobil übertragen werden. Das wären jeden Monat fast 10,8 Exabyte weltweit (vgl. Lobmeier, Sabine 2012).

2.2 Information Overload – die Herausforderung

„Die Zahl derer, die durch zu viele Informationen nicht mehr informiert sind, wächst.“ Rudolf Augstein (1923-2002), dt. Journalist, Herausgeber *Der Spiegel*

Das Zitat des deutschen Journalisten und Gründers des Nachrichtenmagazins *Der Spiegel* verdeutlicht die Kritik am derzeitigen Ist-Zustand der Informationsgesellschaft: Die Konsumenten „leiden“ im Zeitalter der Digitalisierung und Technisierung an Informationsüberlastung, dem sogenannten Information Overload. E-Mails, Blogbeiträge, Bilder, Dokumente und Videos können von den Internetnutzern in dem Bruchteil einer Sekunde in das World Wide Web hochgeladen und abgerufen werden. Und diese Entwicklung beschleunigt sich weiter: Es sind nun nicht mehr allein die Internetnutzer, welche Informationen im Internet produzieren, abrufen und speichern, – in Form von Blogbeiträgen, wissenschaftlichen Arbeiten, in Medienartikeln, bei Facebook, Twitter etc. – sondern auch Milliarden von Geräten, die dasselbe tun: „Smartphones, GPS-Geräte, Kassen, Minicomputer in unseren Autos und anderen Geräten, Browser, die unsere Web-Aktivitäten beobachten, Kameras, Sensoren, Musikgeräte und... und... und...“ (Konitzer 2011). Daten kommen daher aus nahezu unendlich vielen Quellen und sind überall und zu jeder Zeit abrufbar. Dieser Fakt bedeutet jedoch nicht unbedingt, dass die quantitative Steigerung auch eine qualitative Informationszunahme bedeutet. Zu viele Informationen können beim einzelnen Nutzer zu einer Überlastung führen und so kontraproduktiv wirken. Zu diesem Ergebnis kommt auch eine von dem Bundesverband Informationswirtschaft, Telekommunikation und neue Medien (BITKOM) in Auftrag gegebene Studie. Jeder dritte Teilnehmer habe das Gefühl, von Informationen überflutet zu werden (vgl. Scheer 2011). Als Hauptverursacher wird von 71 Prozent der Befragten das Fernsehen genannt, gefolgt vom Internet mit 43 Prozent und dem Radio sowie den Printmedien mit jeweils 24 Prozent.

Als Informationsüberflutung (auch Informationsflut, engl. Information Overload oder Information Flood)

[...] bezeichnet man die Überlastung der am Kommunikationsprozess teilnehmenden Personen mit zum Teil irrelevanten Informationen. Diese Überlastung basiert auf der Tatsache, dass ein Mensch nur eine bestimmte Menge an Informationen während eines bestimmten Zeitraums verarbeiten kann. (Rinne; Rennhak 2006: 2)

Demnach fühlt sich der Rezipient bei einer übergroßen Informationsschwemme zu einem Thema überfordert und ist nicht mehr in der Lage eine Entscheidung zu treffen. Der Ausdruck Information Overload wird gewöhnlich im Zusammenhang mit dem Internet verwendet, Tatsache ist jedoch, dass das Problem keineswegs erst durch das World Wide Web entstanden ist. Bereits 1755 hatte der französische Schriftsteller, Philosoph und Aufklärer Denis Diderot die Idee einer permanenten Informationsüberflutung geäußert und der US-amerikanische Schriftsteller und Futurologe Alvin Toffler den Begriff 1970 in seinem Buch *Future Shock* eingeführt und dadurch popularisiert (vgl. Davis; Shaw 2011: 4). Der amerikanische Internetexperte Clay Shirky geht sogar noch weiter und führt das erste Aufkommen einer Informationsüberflutung auf das Bestehen der Bibliothek von Alexandria zwischen dem 3. und 1. Jh. v. Chr. zurück. Laut Shirky standen der Menschheit zum ersten Mal in der Geschichte mehr Informationen zur Verfügung, als ein einzelner Mensch jemals in seinem Leben hätte verarbeiten können.

But the information overload people are the most narcissistic because information overload started in Alexandria, in the library of Alexandria, right? That was the first example where we have concrete archaeological evidence that there was more information in one place than one human being could deal with in one lifetime, which is almost the definition of information overload. (Juskalian 2008)

2008 hatte Shirky in einem Vortrag auf der Web 2.0 Expo in New York angeführt, dass ein Problem, welches seit so langer Zeit besteht, vielleicht kein Problem mehr ist, sondern ein Faktum (vgl. Sharky 2008: 21:00 bis 23:00 Minute). Das genuine Problem ist also nicht die komplexe Informationsmenge an Daten im World Wide Web, sondern die Mittel, um sie zu bewältigen. Es fehlen demnach Erfahrungen im Umgang mit einer großen Masse an Daten und daraus resultierend herrscht ein Mangel in deren Filterung von Daten. Dies erschwert dem Internetnutzer die Orientierung im Datenmeer (vgl. Konitzer, 2011). Es fehle an der Kompetenz, mit Filterwerkzeugen umzugehen, da geeignete Werkzeuge noch nicht existieren oder nahezu unbekannt sind.

Trotzdem oder gerade deswegen greift nach Aussage von Michael Konitzer, einem deutschen Publizist und Fachmann für digitale Kommunikation, der Begriff Information Overload nicht mehr (vgl. Konitzer 2011). In diesem Zusammenhang sollte der wertfreie Begriff „Big Data“ eingeführt werden, der große Massen an Daten beschreibt, die rasant und unaufhörlich wachsen, „aus vielen verschiedenen Quellen in sehr kurzer Zeit neu entstehen und gleichsam von sehr vielen Anwendern genutzt werden können“ (Landrock; Zilch 2012), sie daher kaum oder schwer beherrschbar scheinen, in diesen polystrukturierten Daten aber ein potentieller Mehrwert liegt. Im Sinne dieser These überfluten die digitalisierten Daten nicht mehr allein den Menschen, sondern helfen Relevanz und Wissen zu generieren, vorausgesetzt man besitzt die dafür notwendige Medienkompetenz, mittels Werkzeugen mit den Daten umzugehen. Mit der Verwendung des neuen Fachterms, würde der negativ konnotierte Begriff des Information Overload dem des wertfreien Big Data weichen und sich neue Chancen ergeben – auch für den Journalismus. Das Marktforschungs- und Analyseunternehmen in der weltweiten Technologie-Industrie Gartner macht daraus eine einfache Gleichung: „Big Data equals Big Opportunity“ (vgl. IBM 2011: 2).

Gleichzeitig ist aus dieser Möglichkeit auch ein Kampf um die Datenhoheit im Netz geworden: Daten avancierten zum Rohstoff, mit dem sich viel Profit erzielen lässt. In diesem Zusammenhang wies der Werbefachmann Michael Palmer darauf hin „Daten sind das neue Öl“ (Palmer 2006). Der Vergleich stimme: Sie sind einer der wichtigsten und ertragreichsten Rohstoffe der heutigen Informationswirtschaft, ähnlich wie es Kohle und Eisenerz in der industriellen Revolution waren (vgl. IBM 2011). Allerdings sind sie nur dann hilfreich, wenn sie entsprechend verarbeitet werden. Unbearbeitete Daten sind wertlos, erst das Interpretieren des Materials und die Rückschlüsse, die daraus gezogen werden, schaffen den gewünschten Mehrwert. „Erst wenn sie bearbeitet würden, entstünden aus ihnen nützliche Produkte, so wie aus Öl Plastik, Dünger oder Benzin gemacht wird“ (Matzat 2011b).

So stellt der immer einfacher werdende Zugang zu globalem Wissen und Informationen, die massenhafte Übertragung und Speicherung von digitalen Daten im World Wide Web sowie der Kampf um die Informationshoheit im Netz den Journalismus vor neue Herausforderungen: Es müssen Strategien erprobt werden, um überhaupt erst an relevante Informationen zu gelangen und nicht im Fluss der Big Data zu versinken. „Denn je größer der Datenberg wird, desto mehr gleicht die Suche nach relevanten Informationen der Suche nach der Nadel im Heuhaufen“ (Bensch 2011b). Journalisten und vor allem Online-Journalisten müssen lernen, sich neue Methoden und Spezialwissen anzueignen, um den Überblick innerhalb der Big Data zu behalten und um ihrer Funktion als „Gatekeeper“ im Datenmeer gerecht zu werden. „Aufgrund der Vielzahl der potentiellen Informationen ergibt sich die schwierige Aufgabe, dass relevante Informationen aus der Menge von Daten gefunden werden müssen“ (Seiermann 2012: 1). Der Umgang mit Daten

führt den Journalisten dazu, sich neue Arbeitstechniken und Werkzeuge anzueignen. Dabei soll die in diesem Sinne neuerworbene Medienkompetenz nicht alleine die Recherche zum Ziel haben. Es soll Wissen produziert werden, damit aus der dispersen Datenmasse Erkenntnisse und Informationen herausgefiltert und dem Rezipienten anschließend verständlich dargestellt werden.

Aus diesen Entwicklungen, den daraus resultierenden Chancen sowie Anforderungen im Umgang mit Daten an den Journalisten von heute, hat sich ein neues Genre im Online-Journalismus etabliert, der Datenjournalismus.

2.3 Definition Datenjournalismus

Bisher herrscht in der Literatur und unter den Fachexperten keine einheitliche Meinung darüber, was genau unter dem Phänomen Datenjournalismus zu verstehen ist. Deshalb erfolgt in diesem Kapitel durch die Darstellung mehrerer umschreibender Definitionen, die den Begriff begrenzen, einordnen und möglichst konkretisieren, eine Annäherung an den Terminus Datenjournalismus.

Datenjournalismus ist eine Form des Online-Journalismus, der auf der Analyse und Filterung großer und vielschichtiger Datenmengen basiert. Er dient dazu, komplexe Zusammenhänge, die in Schrift oder Tabellenform unüberschaubar wären, verständlich aufzubereiten und in genuinen Erzählformen darzustellen (vgl. Dean 2012). Dabei handelt es sich um eine Kombination aus Recherche-Ansatz und neuer interaktiver Veröffentlichungsform (vgl. Matzat 2010d: 6), bei der das Medium World Wide Web zugleich als Recherche- sowie Publikationsplattform dient.

Allerdings arbeitet der Journalismus nicht erst seit Aufkommen dieses neuen Genres mit elektronischen Daten. Seit der flächendeckenden Einführung von Computern und Internetzugang zu Beginn des neuen Jahrtausends bedienen sich klassische journalistische Medien sowie der Online-Journalismus des Internets als Rechercheinstrument und unterfüttern ihre Berichterstattung mit Statistiken, Fotos und Infografiken (vgl. Schneider; Raue ⁵2003: 251 ff.). Das Neue am Datenjournalismus ist laut Christina Elmer, Datenjournalistin beim Wochenmagazin *Stern*, dass der Datensatz Ausgangsbasis und Gegenstand der Berichterstattung zugleich ist: „In der neuen Form geht es anders herum, dass ich mir zuerst die Daten anschau und dann in der Auswertung eine Idee davon bekomme, was ist eigentlich meine Geschichte“ (Zapp 2011: 0:40 bis 0:55 Minute). So dienen laut Elmer Daten nicht mehr allein zur Unterfütterung einer vorangestellten These, sondern erst aus ihnen entsteht die eigentliche Fragestellung. Wurden Daten im klassischen Journalismus vorher nur zur Ergänzung von Beiträgen zum besseren Verständnis des Inhalts genutzt, verwendet der Datenjournalismus die unverarbeiteten Rohdaten nicht mehr rein als Recherchequelle, sondern erhebt sie zur Grundlage jeglicher Berichterstattung und zentralem Gegen-

stand der Geschichte und Präsentation (vgl. Matzat 2011c). So erklärt auch der in der politischen Bildung und als Medienpädagoge tätige Gründer des *Open Data Network e.V.* und Journalist Lorenz Matzat die Entwicklung zum heutigen Datenjournalismus: „[...] Das Neue daran ist, Datensätze zum Gegenstand des Berichts und zu einer nützlichen interaktiven Rechercheumgebung zu machen. Bislang sind Datensätze oft als Beiwerk oder Unterstützung gesehen worden, jetzt ist der Datensatz an sich der Gegenstand der Berichterstattung“ (Medium Magazin 2011). Vereinfacht lässt sich dieser Sachverhalt laut der Deutschen Presse-Agentur (vgl. Elmer 2012: 3,4) in folgender Gleichung darstellen:

1. Ursprüngliche Form: These → Daten

Funktion der Daten: Beweismittel, Einordnung, Relevanz verdeutlichen

2. Datenjournalismus: Daten → These

Funktion der Daten: Dienen als Fundgrube für inhaltlichen Mehrwert

Der hier definierte Datenjournalismus existiert nur im und lebt von den Möglichkeiten des World Wide Web. Allein durch die Mittel und Optionen, die das neue Medium bietet, erhält der Datenjournalismus seine eigentliche Form. Für den britischen Online-Journalisten, Blogger und Professor an der *City University's School of Journalism* in London, Paul Bradshaw kommt er deshalb – rein faktisch – nur online vor. Er begründet das so:

Data journalism may be a new term, but the idea of ploughing through information to find the story is as old as journalism itself – the difference is that the internet and computing power are giving us new tools to help at every stage of the journalistic process. (Bradshaw 2011: 49)

Weiter erläutert Bradshaw, dass egal ob es nun in erster Linie um die Suche nach Daten, deren Abfrage, Visualisierung oder deren Kombination zu einer sinnvollen Geschichte im World Wide Web geht, der Datenjournalismus ein wesentliches Merkmal für alle journalistischen Publikationsplattformen ist (und deren Zukunft dominieren wird), er seinen Höhepunkt nur online erreichen kann (vgl. Bradshaw 2011: 49 ff.). Auch das European Journalism Centre hielt während einer DDJ-Journalisten-Gesprächsrunde im August 2010 den zwingenden Faktor „online“ des Datenjournalismus in einem Auszug fest. Dort beschreibt der Zirkel von Fachexperten, dass die Nutzung von Daten als Grundlage der Berichterstattung vor zehn Jahren noch schwierig und kostspielig war, da die IT-Kompetenzen zur Recherche und Darstellung von großen Datenmengen in den Medien nicht ausreichend vorhanden waren. Datenbanken wurden demnach hauptsächlich durch investigative Jour-

nalisten genutzt. Redakteure und Reporter verließen sich in der Regel auf Informationen und aufbereitete Daten von externen Quellen. Der Online-Datenjournalismus veränderte diese Sicht:

Today there is a notable change. Collections of data are becoming available online, often for free. There is a whole stack of tools for digging into “big data”. Open source tools in the World Wide Web allow navigation and analysis of large amounts of data rather quickly. (European Journalism Centre 2010: 6)

Der Blogger Moritz Shane umschreibt diesen Zustand während der Diskussion über die Veröffentlichung des *Data Journalism Handbook* im April 2012 folgendermaßen: Datenjournalismus sei eine Kombination von klassischem Journalismus mit den neuen Mitteln, die durch das Vorhandensein und die Zugänglichkeit riesiger Datenmengen im World Wide Web und deren „[...] Möglichkeiten der computer- und Web-2.0-unterstützten Analyse und Darstellung dieser Daten entstanden sind“ (Beckedahl 2012). Stefan Plöchinger, Journalist bei süddeutsche.de, bringt diesen Fakt auf den Punkt: „Sie ist deshalb so besonders, weil sie nur mit Computertechnik und den Möglichkeiten des Internets denkbar ist. [...] Erst die digitale Welt erschließt dem Journalismus solche ungekannten Präsentationsmöglichkeiten“ (Plöchinger 2012b).

Der Begriff Datenjournalismus kommt ursprünglich aus dem Englischen, man spricht hier vom “data-driven journalism” kurz “DDJ”. Wörtlich übersetzt bedeutet dieser Name „datengetriebener Journalismus“ und setzt damit voraus, dass die Berichterstattung nur mit Unterstützung von Datensätzen geschieht; „[...] im Internet per interaktiven Anwendungen, die auf Datenbanken zugreifen“ (Matzat 2011c). Das Wort selbst beinhaltet jedoch auch den Begriff „Journalismus“ und zeigt damit, dass der Fokus trotz des verstärkten Dateneinsatzes letztendlich auf der journalistischen Aufbereitung liegt (vgl. Dean 2012). Neben Datenjournalismus wird im Deutschen auch häufig der Begriff „Datenbankjournalismus“ oder „Datensatzjournalismus“ verwendet. Im Englischen trifft man auf Synonyme wie “computational journalism”, “visual journalism” sowie “databank journalism” oder schlicht “datajournalism” (vgl. Matzat 2011c).

Manche Datenjournalisten, wie zum Beispiel Lorenz Matzat, betrachten die computergestützte Recherche als direkten Vorläufer des Datenjournalismus (vgl. Matzat 2011c). Andere wiederum sehen das CAR als Teil des Datenjournalismus (vgl. Focke 2011); als Arbeitsmethode, die Anwendung findet auf dem Weg zum fertigen datenjournalistischen Produkt. Unumstritten ist jedoch eine direkte Verbindung zwischen den beiden Genres.

Das CAR wird im englischsprachigen Raum seit den Sechziger Jahren praktiziert und von zahlreiche Universitäten und Journalistenverbänden gelehrt (vgl. Hartmann 2011). Man versteht heutzutage darunter eine Technik, die große Datenmengen unter der Verwendung eines Computers sammelt, analysiert und aufbereitet. Dabei kommen Statistikprogramme

und/oder Informationssysteme zum Einsatz. Laut Christina Elmer geht es beim CAR darum, mithilfe von Computern im Netz nach bestimmten Daten zu suchen. „Das geht von der einfachen Google-Suche bis zu aufwändig programmierter Hacker- und Robotersoftware, die auch Daten alter Internetseiten aufspüren kann“ (Focke 2011). Die computergestützte Recherche wird vorrangig zur investigativen Recherche genutzt, um Auffälligkeiten und Unregelmäßigkeiten in Datensätzen zu entdecken, die Anlass für eine tiefergehende Recherche bieten könnten (vgl. Krampe 2012d). Dabei grenzt CAR sich vom Datenjournalismus in der Form ab, dass CAR rein auf Seiten der Recherche passiert, während beim Datenjournalismus die Datensätze nicht nur als Recherchequelle fungieren, sondern die Daten zum zentralen Gegenstand der Geschichte und deren Präsentation erhoben werden (vgl. Matzat 2011c).

So gehören zu der konsequenten Umsetzung von datenjournalistischen Projekten neben der Recherche, dem Analysieren und Auswerten der Datensätze und deren anschließenden verständlichen Präsentation durch Visualisierungen, auch immer das Bereitstellen der Rohdaten – auf denen das Werk fußt – für den Nutzer (vgl. Langer 2011b: 4). Dadurch gewährleistet der Journalist größtmögliche Transparenz seiner Arbeit und verfolgt den Grundsatz von Open Data, der eine wichtige Voraussetzung für guten Datenjournalismus darstellt (vgl. Kaute 2011). Die Open-Data-Bewegung setzt sich dafür ein, dass Behörden möglichst viele ihrer Datensätze, die keiner Geheimhaltung unterliegen, in offenen, maschinenlesbaren Formaten zur Verfügung stellen (vgl. Matzat 2010c). Im Kapitel 3.1.1 wird näher auf diesen Terminus und seine Relationen zum Datenjournalismus eingegangen.

Während der konventionelle Journalismus darauf abzielt, Beiträge in Form von klassischen journalistischen Darstellungsformen zu publizieren, nutzt der Datenjournalismus neue interaktive Erzählformate zur Veröffentlichung seiner gewonnenen Ergebnisse, wie zum Beispiel Datenbanken, Datenvisualisierungen oder datengestützte Anwendungen für das Netz und mobile Endgeräte (vgl. Heidmeier 2012). Im Datenjournalismus wird Wissen aus anderen Fachbereichen genutzt wie zum Beispiel über Statistik, Programmierung, Visualisierung (Design) und Storytelling (vgl. Ausserhofer 2010: 4) und kombiniert diese Aspekte mit investigativen Rechercheprozessen (vgl. Bradshaw 2010).

Diese daraus entstandenen, neuen datenjournalistischen Erzählformen führen zugleich zu mehr Qualität im (Online-)Journalismus. So ermöglicht der DDJ interaktive Rechercheumgebungen in großen Datenbeständen, die dem Nutzer erlauben, selbst aktiv mit den Daten umzugehen und Informationen in den Daten zu entdecken.

Die Nutzer/innen – ob Journalist/innen oder Leser/innen – können nach individuellen Interessen in den Datenbeständen recherchieren oder Vorgänge beobachten (monitoren). Und entgegen des klassischen Datenjournalismus kann der moderne durch leistungsfähige Software

auch komplizierte gesellschaftliche Vorgänge und Zusammenhänge aufbereiten und on-demand abbilden. (Matzat 2010b)

So bekomme der Nutzer die Möglichkeit, in Eigenarbeit durch Zahlenberge zu navigieren, sich Daten auf eigene Weise zu erschließen und dadurch selbst zu entscheiden, wie tief er in das Thema einsteigen möchte. Ein Beispiel dafür sind die vom *The Guardian* datenjournalistisch, in interaktiven Anwendungen aufbereiteten Warlogs aus dem Jahr 2010. „Sie überlassen dem User, ob er einfach nur Art und Umfang von Kampfhandlungen in einer interaktiven Grafik nachvollziehen will oder die Logs im Detail liest“ (Leßmöllmann 2012). Ein weiterer entscheidender Faktor datenjournalistischer Projekte kann die Einbindung des Crowdsourcing-Prinzips sein. Der Prozess beschreibt die Einbindung einer Gruppe von Internetnutzern (die sich außerhalb der Strukturen des publizierenden Mediums befindet) in die Recherche und/oder Produktion von Inhalten (vgl. Pelzer 2011). So werden zum Beispiel für das Projekt „Zugmonitor“ der *Süddeutschen Zeitung* die Leser öffentlich dazu aufgerufen, selbst aktiv am Projekt teilzunehmen, zu recherchieren und zu analysieren (vgl. Plöchinger 2012b). Letztlich meint Crowdsourcing im Datenjournalismus jedoch nicht, dass Laien die Rollen der Profis übernehmen, sondern diese bei ihrer Arbeit unterstützen. „Die Bewertung, Einordnung und Veröffentlichung bleibt üblicherweise im Verantwortungsbereich einer Redaktion“ (Forum Journalismus und Medien Wien 2012).

Auf die genaue Differenzierung, Anwendungsbereiche und Zwecke der verschiedenen datenjournalistischen Erzählformen wird in Kapitel 3.3 genauer eingegangen.

2.3.1 Charakteristika

So sehr sich die diversen Definitionsversuche des Datenjournalismus in ihren Ausführungen unterscheiden, so stimmen sie jedoch alle im Kern überein, wenn es um die Aspekte geht, die einen Beitrag zu einem datenjournalistischen Stück werden lassen. Diese Grundcharakteristika können dabei als Gerüst und Voraussetzung für den Datenjournalismus angesehen werden, „um das dann jegliche weitere Schritte gesponnen werden können“ (Krampe 2012a). Im Folgenden werden neun Merkmale für datenjournalistische Werke erläutert, die sich immer wieder in relevanten Texten über Datenjournalismus finden lassen und sich aus den vorherigen Definitionen (siehe Kapitel 2.3) ergeben.

- Daten liegen in strukturierter Form digital vor und können digital weiterverarbeitet werden (vgl. Matzat 2010a).
- Der Datensatz ist Ausgangsbasis und Gegenstand der Berichterstattung zugleich (vgl. Bradshaw 2010).
- „Die Daten stehen im Mittelpunkt einer investigativen Recherche“ (Langer 2011b: 5).

- Es geht um öffentliche, nicht um private Daten (vgl. Langer 2011b: 5; siehe Kapitel 3.1.1).
- Ein vorhandener, maschinenlesbarer Datensatz wird per Software analysiert und miteinander verknüpft. Gegebenenfalls werden auch mehrere maschinenlesbare Datensätze miteinander verbunden (engl. mashup) (vgl. Krampe 2012a).
- Obligatorisch ist die anschließende Darstellung der Erkenntnisse, nicht die reine Recherche. Dabei ist die multimediale und/oder interaktive Darstellung der Daten in einer Anwendung (vgl. Krampe 2012a), folglich die Aufbereitung in interaktiven Formaten im World Wide Web obligatorisch. In welcher Form diese geschieht, ob durch Datenbanken, visuelle Grafiken und/oder Animationen (siehe Kapitel Erzählformen), ist zurzeit noch nicht klar umrissen (vgl. Krampe 2012a).
- Der Leser kann selbst aktiv mit den Daten umgehen, sie sich individuell erschließen und Vorgänge beobachten (monitoring) (vgl. Matzat 2010b).
- Die Journalisten veröffentlichen den ganzen Datensatz, also nicht nur die für ihren Beitrag genutzten Daten, so dass die Rohdaten auch von anderen weiter genutzt werden können und die Recherche gläsern, d.h. transparent wird (vgl. Krampe 2012a).
- Die Nutzer können in das datenjournalistische Projekt miteinbezogen und zur aktiven Mitarbeit aufgerufen werden (vgl. Open Data Network 2010: 5; siehe Kapitel 3.3.4).

Die Journalistin und freie Mitarbeiterin bei der Wochenzeitung *Der Freitag*, Anna-Lena Krampe, hat diese Punkte grafisch wie folgt umgesetzt:

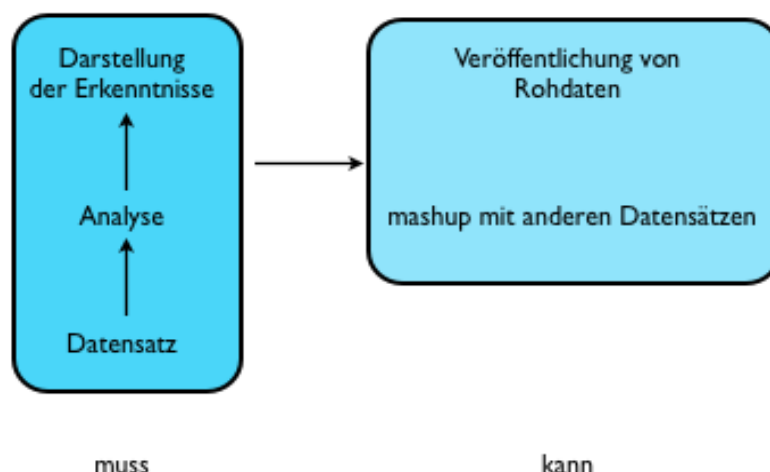


Abbildung 1: Faktoren zur Kennzeichnung einer datenjournalistischen Arbeit. Quelle: Krampe 2012c.

2.3.2 Meilensteine des modernen Datenjournalismus

Als einer der Pioniere des modernen Datenjournalismus wird Adrian Holovaty angesehen (vgl. Bradshaw 2011: 49). Er ist Journalist und Webprogrammierer aus den USA und hat 2007 die hyperlokale Plattform *Everyblock* aufgebaut. 2006 schrieb Holovaty den für den Datenjournalismus wegweisenden Text „A fundamental way newspaper sites need to change“. In seinem Essay beschreibt der Amerikaner, was sich an den Internetauftritten von Zeitungen alles verändern müsse, damit diese, beispielsweise gegenüber Blogs, zukünftig konkurrenzfähig bleiben könnten. Journalisten stünden vor der großen Herausforderung, ihre „auf Geschichten basierende Sichtweise“ (engl. storybased telling) abzulegen und ihre Aufgabe im Sammeln von Informationen und dem anschließenden Verfassen eines Zeitungsartikels zu sehen (vgl. Krampe 2012c). Weiter sagt er, dass die Zielsetzung bei dieser Art von Journalismus sein müsse, „wichtige, konzentrierte Information, die für das Leben der Menschen nützlich ist und ihnen hilft, die Welt zu verstehen“ (Mahmood 2011), zu liefern.

So much of what local journalists collect day-to-day is structured information: the type of information that can be sliced-and-diced, in an automated fashion, by computers. Yet the information gets distilled into a big blob of text – a newspaper story – that has no chance of being repurposed. (Holovaty 2006)

Laut diesem Zitat müssen journalistische Produkte zu einem neuen, zu einem anderen als dem ursprünglichen Zweck wiederverwendet (repurpose) werden können (vgl. Krampe 2012c). Denn in Rohdaten stecke nach Holovaty (vgl. Holovaty 2006) meistens viel mehr als die Information, die ein Journalist für einen einzigen Beitrag auswähle. Laut Matzat (vgl. Matzat 2011c) lasse sich so auf Dauer Mehrwert abschöpfen. „Denn in diesen strukturierten Informationen lässt sich auch nach Jahren noch einfach recherchieren; verknüpft mit anderen Datenbanken können publizistische Angebote damit angereichert werden“ (Matzat 2011c).

Im Weiteren führt Holovaty einige Beispiele aus dem Alltag an, die wichtige Rohdaten liefern. Würden zum Beispiel bei einem Feuerwehreinsatz die Daten zur Zeit und Art des Brandes, zum Ort des Geschehens, den Opfern oder auch der Entfernung des Ortes zur nächsten Feuerwehrrache dokumentiert werden, könnte daraus im Laufe der Zeit eine Datenbank konstruiert werden, die viel über die Einsätze und die Arbeit der Feuerwehr verraten könne (vgl. Krampe 2012c). Laut Holovaty könnten die Daten so systematisch gesammelt und mit vergangenen oder auch künftigen Einsätzen verglichen werden: „That’s what I mean by structured data: information with attributes that are consistent across a domain. Every fire has those attributes, just as every reported crime has many attributes, just as every college basketball game has many attributes“ (Holovaty 2006).

Laut Matzat hat die Zeit Holovaty Recht gegeben: „data-driven-journalism oder Datenjournalismus hat seit 2009 eine steile Karriere hingelegt“ (Matzat 2011c). So unterhalten *The Guardian* und die *New York Times* mit sogenannten Datastores und Data-Desks seit 2009 eigene Ressorts, die sich um das Aggregieren, Gewichten und Auswerten von Daten für redaktionelle Beiträge kümmern (vgl. Dean 2012). Datenjournalismus-Projekte sind in den Medien der USA und England fester Bestandteil der Berichterstattung. So stellte die britische Tageszeitung *The Guardian* im Frühjahr des selben Jahres 485.000 Dokumente über den Spesenskandal des Houses of Parliament ins Netz und bat die Bürger um aktive Mitarbeit, das Material der Rechnungslegung der Parlamentarier zu analysieren und Einzelheiten zu den Aussagen der Abgeordneten herauszufinden (vgl. Ausserhofer 2010: 26). „Wir gaben unseren Benutzern die Möglichkeit, die Datenbank mit den Ausgaben aus der Staatskasse einzusehen – wir veröffentlichten die Daten hinter den Meldungen“ (Rogers 2012). Noch nie zuvor in der Geschichte des Online-Journalismus wurde eine derart umfangreiche Berichterstattung auf Basis eines Datensatzes durch die Mithilfe der Öffentlichkeit über das World Wide Web ermöglicht und dies blieb nicht ohne Konsequenz: Es folgten noch im selben Jahr Rücktritte und Gesetzesänderungen in Großbritannien.

Doch die eigentliche Sternstunde des Datenjournalismus brach erst im Folgejahr an: Im Sommer 2010 veröffentlichte die Internetplattform *Wikileaks* die sogenannten Warlogs (Kriegstagebücher) der US-Armee aus Afghanistan. Hunderttausende geheimer Dokumente publizierte die Whistleblowing-Webseite und machte die Daten allgemein frei verfügbar. Neben *The Guardian* und der *New York Times*, beteiligte sich auch *Der Spiegel* an der Bearbeitung und Präsentation der Datenberge. Laut Simon Rogers ist „die Veröffentlichung der Wikileaks-Daten zum Afghanistankrieg [...] ein Triumph des investigativen, an Fakten orientierten Datenjournalismus [...]“ (Rogers 2011: 119) und Matzat konstatiert, dass selten zuvor ein Datensatz solch großes mediales Interesse auf sich zog (vgl. Matzat 2011c). Jedoch waren die Dokumente „in den seltensten Fällen in erzählerischem Ton verfasst, der von Journalisten in größeren Stories hätte zitiert werden können“ (Rogers 2012). Vielmehr handelte es sich um knappe technische Informationen über Vorgänge, Zwischenfälle, Anschläge mit präziser Orts- und Zeitangabe (vgl. Rogers 2012). Erst der Datenjournalismus, der mit seinen Methoden die riesigen Informationsberge mit Statistikwerkzeugen durchdrang, sie anschließend visuell veranschaulichte und dadurch das vorherige Rohmaterial für journalistische Zwecke brauchbar machte, schaffte einen für den Leser verständlichen Zugang zum Material. So benutzte *The Guardian* die von 90.000 militärischen Einsatzprotokollen aus Afghanistan von Julian Assanges Whistleblower-Portal nicht nur für herkömmliche und multimedial angereicherten Berichte, sondern erstellte auch eine für Nutzer in mehreren Ebenen recherchierbare Datenbank – ein Vorgehen, das international hoch gelobt wurde (vgl. Langer 2011a: 1). Es entstand eine neue Form des

publizistischen Erzählens: „Jenseits von Text, Ton und Bild entsteht so mit dem Datenjournalismus ein neues interaktives Erzählformat in Zusammenspiel mit Datenbanken“ (Matzat 211b). Es folgten im Oktober die Aufbereitung der *Iraq WarLogs* und der geheimen Depeschen der US-Botschaften im Dezember, dessen Rohmaterial ebenfalls von *Wikileaks* publiziert wurde. Während sich bei den Afghanistan-Tagebüchern nur *The Guardian* des DDJ erfolgreich bedient hatte, versuchten nun auch diverse andere Medien, die Datenberge durch die Methoden des DDJ anschaulich und verständlich umzusetzen (vgl. Beckedahl 2010). Teilweise entstand daraus eine datenjournalistische Kooperation internationaler Medien.

Seit diesem Zeitpunkt schenkt der Journalismus, insbesondere der Online-Journalismus, dem DDJ immer mehr Beachtung. Viele marktführende Online-Medien haben das Potential des Datenjournalismus erkannt und nutzen seit 2010 gehäuft Daten als Grundlage für ihre publizistischen Werke. So stieg die Zahl datenjournalistischer Projekte im Internet in den letzten Jahren stetig an. „In ein paar Jahren ist der Datenjournalismus vielleicht schon keine Neuerung mehr; heute ist er es noch. Die Welt ändert sich, und sie ändert sich durch Daten“ (Rogers 2012).

2.3.3 Journalistische Einordnung des Datenjournalismus

Klaus Meier, Professor am Studiengang Journalistik der Katholischen Universität Eichstätt-Ingolstadt, ordnet den hier dargestellten Gegenstand des Datenjournalismus zum Typus des Präzisionsjournalismus (vgl. Meier 2011). Diesen Sachverhalt führt Meier auf den Österreicher Philip Meyer zurück, der den Begriff „Präzisionsjournalismus“ bereits 1973 in seinem Buch *Precision Journalism* etablierte. Laut Meyer sollen sich Journalisten bei der Recherche zu aktuellen Themen der Methoden der empirischen Sozialforschung bedienen, um zu wissenschaftlich abgesicherten Fakten über soziale Sachverhalte zu gelangen (vgl. Kunczik 2001: 165) und dadurch präziser und vor allem „objektiver“ berichten zu können. Präzisionsjournalismus ist nach der Definition von Meyer auch immer wissenschaftlicher Journalismus: „It means treating journalism as if it were a science, adapting scientific method, scientific objectivity, and scientific ideals to the entire process of mass communication“ (Meyer 2012). Laut Meier fußt das neue Genre Datenjournalismus auf diesem Ansatz und lässt sich daher wissenschaftlich diesem zuordnen. Dabei verschiebe sich das klassische Rollenbild des Journalisten im DDJ: War der Nachrichtenerstatter vorher der „Vermittler“ (objektiver Journalismus) oder „Wachhund“ (investigativer Journalismus), wird er im neuen interdisziplinären Feld des Datenjournalismus zum „Forscher“ (vgl. Meier 2011).

Auch der Datenjournalist Matzat sieht die Relation zwischen den beiden Journalismusformen, geht jedoch in der wissenschaftlichen Einordnung des Datenjournalismus noch einen Schritt weiter. Neben der Ansiedlung im Bereich des Präzisionsjournalismus sieht Matzat den Datenjournalismus als interdisziplinäres Feld, der sich in drei Bereichen überlagert: „erstens visu-

eller Journalismus oder Infografiken, zweitens multimediales und interaktives Storytelling und drittens investigativer Journalismus“ (Matzat 2011c). Zum besseren Verständnis werden die beiden Begriffe „Infografik“ und „Multimedia-Storytelling“, wie sie in diesem Zusammenhang zu verstehen sind, wie folgt erläutert.

Die Infografik ist klassifiziert als journalistische Darstellungsform, die während der Achtzigerjahre aus einer Verschmelzung der beiden Worte Information und Grafik entstand (Bouchon 2006: 11-15). Dabei sind „Infografiken [...] gezeichnete Nachrichten oder Erläuterungen von komplizierten Sachverhalten“ (Walther 2001: 4), deren Ziel es ist, Informationen grafisch aufzubereiten sowie zu präsentieren. Dabei dient sie nicht rein künstlerischen oder dekorativen Zwecken, denn schon das im Begriff enthaltene Wort „Information“ (lat. informatio = Bildung, Belehrung) bestimmt, dass Infografiken Neuigkeiten mit nachrichtlichem Wert vermitteln (vgl. Haller 2000: 213).

Eine Multimedia-Story ist laut der Knight Foundation der US-Universität Berkeley (vgl. Stevens 2011) eine Kombination aus Text, Fotos, Videoclips, Audios und/oder Grafiken auf einer Website, die in einem nichtlinearen Erzählformat interaktiv dargestellt werden. Dabei ergänzen sich die Informationen aus den verschiedenen Medienquellen und wirken komplementär, anstatt sich zu doppeln. In diesem Sinne bedeutet non-lineare Rezeption, dass anstatt eines einzigen streng-strukturierten Erzählweges der Benutzer wählen kann, zu welchen Elementen einer Geschichte er navigiert.

Unter investigativem Journalismus wird eine Form des Journalismus verstanden, bei der durch intensive Recherche bisher der Öffentlichkeit vorenthaltene oder verschwiegene Informationen von politischer oder gesellschaftlicher Relevanz bekannt gemacht werden (vgl. Posch 2009: 63). „Er möchte entweder offiziell gedeckte oder tatsächlich (noch) nicht bekannte Missstände in Politik, Wirtschaft, Kultur und Gesellschaft ans Licht der Öffentlichkeit bringen“ (Kunczik 2001: 163)

Betrachtet man diesen Sachverhalt aus der Perspektive der Journalismusforschung, so liegt der DDJ in der Schnittmenge zwischen einer Darstellungsform (Infografiken), einem Kommunikationsmodus (Multimedia-Storytelling) und dem Berichterstattungsmuster des investigativen Journalismus (vgl. Lorenz-Meyer 2011).

3 DATEN IM NETZ

3.1 Datengewinnung

Grundlage jeglicher Berichterstattung und zentraler Gegenstand für alle datenjournalistischen Werke ist das Vorhandensein eines Datensatzes; Der „Datenjournalismus lebt von Daten“ (Grieß 2011). So ist laut Simon Rogers, leitender Redakteur des Datablogs des britischen *The Guardian* „der erste Schritt des Datenjournalisten [...], Daten überhaupt als interessante Quelle zu erkennen, aufzuspüren und ‚journalistisch urbar‘ zu machen“ (Leßmöllmann 2012). Das Web steckt ihm zufolge voller Daten, die nur darauf warten, von Journalisten aufgespürt und sinnvoll verarbeitet zu werden. Dabei stehen Journalisten durch die Digitalisierung und das World Wide Web heute so viele Informationen zur Verfügung wie noch nie zuvor. Doch welche Daten sind in diesem Meer aus Zahlen für den DDJ relevant und welche Wege nutzen Journalisten, an verwertbaren Datensätze und Rohdaten zu gelangen? Lassen diese sich für bestimmte Zwecke sinnvoll kategorisieren? Zur Beantwortung dieser Fragen werden in den folgenden Kapiteln drei zentrale Typen von Daten im Netz vorgestellt: Zum einen werden offen gelegte Daten aus öffentlicher Hand, aus der Wissenschaft, von privaten Wirtschaftsunternehmen, Nichtregierungsorganisationen (NGOs), oder Privatpersonen, hier beschrieben als Open Data (1), so wie das Prinzip des Aussagens von Webseiten, das sogenannte Webscraping (2) und die dadurch gewonnenen Transparent Data näher erläutert. Außerdem wird auf Informationen vom Whistleblowing, den Leaked Data (3) und die Möglichkeiten dieser Datensätze für den DDJ eingegangen.

Aus diesen Datenformen ergibt sich für Andreas Grieß zwangsläufig das Modell der Datenzwiebel (siehe Abb. 2). Dabei stelle jede Schicht eine andere Datenmenge und Entwicklung dar. „Umso weiter man nach innen dringt, umso interessanter werden die Daten in der Regel“ (Grieß 2011). Für den in dieser Arbeit dargestellten Datenjournalismus wird das Modell jedoch mit insgesamt drei statt fünf Schichten beschrieben.

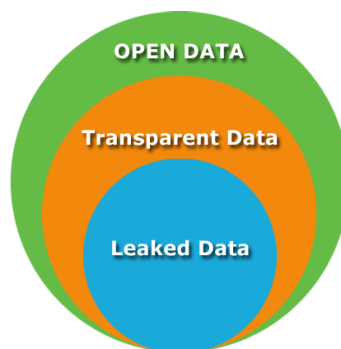


Abbildung 2: Die Datenzwiebel. Eigene Grafik nach Grieß 2011.

3.1.1 Das Prinzip Open Data

Statistiken, medizinische Forschungsergebnisse, Lehrmaterialien, Geodaten, Umweltinformationen und Projektevaluationen: Regierungen, Universitäten, Organisationen, Stiftungen und Wirtschaftsunternehmen sammeln und produzieren riesige Datenmengen, um ihre Aufgaben zu erfüllen (vgl. Betterplace Lab 2012). Bislang ist eine Vielzahl dieser Daten unter Verschluss und so noch nicht für die Allgemeinheit zugänglich. „Das soll sich im Zuge der Open-Data-Bewegung ändern“ (Betterplace Lab 2012), liegt doch ein immenses Potential in ihnen, das ungenutzt gelagert und für deren Analyse der Staat oder Unternehmen keine Kapazität haben, auf dessen Basis neues Wissen, Dienstleistungen und Produkte erarbeitet werden könnten. Die Open-Data-Bewegung möchte diesen Zustand ändern.

Die Open-Data-Bewegung ist eine Kultur, die sich mit dem Umgang von digitalen Daten beschäftigt. Ihr Ziel ist es, Daten und Informationen zu öffnen und für jeden zugänglich zu machen, folglich mehr Transparenz und Freiheit für digitalisiertes Wissen zu schaffen. Dabei sollten die Datensätzen in strukturierter, maschinenlesbarer Form oder über eine offene Schnittstelle (API) im World Wide Web weiterverarbeitet und weiterverbreitet werden können (vgl. Paderta 2012). In Europa sind derzeit zwei Stiftungen für den Diskurs um Open Data richtungsweisend: die *Sunlight Foundation* und die *Open Knowledge Foundation*. „Deren Definitionen zu freien Daten und Informationen dienen als Richtschnur im Open Data-Prozess“ (Paderta 2012). Die *Open Knowledge Foundation* hat den Begriff Open Data in ihrem online publizierten *Open Data Handbook* folgendermaßen definiert: „Offene Daten sind Daten, die von jedermann frei benutzt, weiterverwendet und geteilt werden können – die einzige Einschränkung betrifft die Verpflichtung zur Nennung des Urhebers“ (Open Knowledge Foundation 2012). Weiter erklären die Open-Data-Aktivist*innen in ihrem Handbuch, dass Open Data auf drei verschiedenen Ebenen, einer technischen, einer rechtlichen und einer soziokulturellen, verläuft. Die technische umfasst dabei die Bereitstellung (Format, Metadaten, Standards), die rechtliche Ebene enthält Lizenzen, Richtlinien und Gesetze, „die soziokulturelle Dimension beinhaltet das Verständnis des Staates und des Umgangs mit immateriellen Gütern in sozialen Prozessen“ (Paderta 2012).

Das Open-Data-Prinzip ist im wissenschaftlichen Bereich nichts Neues und ähnelt dabei vielen anderen „Open“-Konzepten wie Open Access (freier Zugang zu wissenschaftlichen Veröffentlichungen (vgl. Diedrich 2011)), Open Content (freier Zugang/Anwendung, uneingeschränkte Modifikation und Weiterverbreitung von Texten, Bild- und Tonwerken bei bestehendem urheberrechtlichem Schutz durch freie Lizenzen; geprägt durch die Einführung der Creative-Commons-Lizenz (vgl. Definition of Free Cultural Works 2012)) und Open Source (offener Quellcode und Software, die jeder nach Belieben studieren, benutzen, verändern und kopieren darf (vgl. Bundeszentrale für politische Bildung 2012)). Die Open-Data-Initiative befindet sich jedoch im Gegensatz zu der fast 30 Jahre alten Open-Source-Software-

bzw. der Free-Software-Bewegung und der Open-Access-Initiative, welche ihren Anfang bereits vor Beginn des 21. Jahrhunderts hatte, noch am Anfang ihrer Entwicklung (vgl. Paderta 2012).

Unter „offenen Daten“ (englisch Open Data) werden laut *Open Knowledge Foundation* alle Datentypen subsumiert, die ohne jegliche Einschränkung zur Weiterverwendung und -bearbeitung für jeden frei zugänglich sind. Das Konzept umfasst dabei nicht nur die Daten der Regierung und der öffentlichen Verwaltung (vgl. Open Knowledge Foundation 2012). Laut der Internet & Gesellschaft Co:laboratory geht es bei Open Data auch um Daten aus Bibliotheken und Archiven, von privatwirtschaftlich agierenden Unternehmen, der Wissenschaft sowie Non-Profit-Einrichtungen (vgl. Internet & Gesellschaft Co:laboratory 2010: 27). Wichtig ist in diesem Kontext, dass es sich bei den Daten ausschließlich um nicht personenbezogene oder weitestgehend anonymisierte Daten handelt, also um Daten, die keinen Datenschutz- oder Sicherheitsbeschränkungen unterliegen. Das Open-Data-Prinzip und der DDJ verfolgen somit beide denselben Grundgedanken im Gebrauch von Daten (siehe Kapitel 2.3.1 Charakteristika); sie wollen mehr Transparenz durch den offenen Umgang mit Rohdaten und deren Bereitstellung als eine für alle frei zugängliche Informationsquelle. Open Data ist somit neben anderen Faktoren eine wichtige Voraussetzung für guten Datenjournalismus (vgl. Kaute 2011).

Für Open Data wird heute oft der Begriff Open Government Data synonym verwendet, „also „offene Daten der öffentlichen Verwaltung“ oder kurz „offene Behördendaten“. Offene Behördendaten spielen eine wichtige Rolle im Prozess der Öffnung von Regierung und Verwaltung, der als Open Government bezeichnet wird“ (Dietrich 2011). Dabei meint Open Government Data alle Datenbestände des Staates, „die im allgemeinen Interesse der Öffentlichkeit ohne Einschränkung zur freien Nutzung, zur Weiterverbreitung und zur freien Weiterverwendung frei zugänglich gemacht werden“ (Paderta 2012) und deren Erstellung mit öffentlichen Mitteln finanziert wurde. Dies schließt gemäß des Open-Data-Prinzips alle Daten des öffentlichen Sektors aus, die Personenbezogen sind, die nicht im Bereich des öffentlichen Interesses publiziert werden, die der Geheimhaltungspflicht unterliegen, „sowie Betriebs - und Geschäftsgeheimnisse beinhalten, so dass einer Freigabe ohne Rücksprache mit den Betroffenen nicht statt gegeben werden darf“ (Zeppelin Universität 2010). Es geht im engeren Sinne um Statistiken, Geodaten, Karten, Pläne, Umwelt - und Wetterdaten, Materialien der Parlamente, Ministerien und Behörden, Haushaltsdaten, Gesetze, Verordnungen, Satzungen, richterliche Entscheidungen und sonstige Veröffentlichungen (vgl. Zeppelin Universität 2010). Wenn diese vom Staat und der Regierung bereitgestellten Daten strukturiert und maschinenlesbar öffentlich zugänglich gemacht werden, können (Daten-)Journalisten sie durchsehen, durchsuchen, filtern, aufbereiten und weiterverarbeiten. Diese im Rahmen der Open-Data-Initiative frei gegebenen, staatlichen Datensätze sind somit eine der wichtigsten Quellen für den Datenjournalismus und sei-

ne datengestützte Berichterstattung. Derzeit sind viele der geforderten Daten bereits online zugänglich, oft sind diese jedoch nicht maschinenlesbar, da sie in Form von Textdokumenten in proprietären Formaten vorliegen. Zudem sind die Daten an zu vielen Stellen verteilt und schwer auffindbar. Oft fehlt sogar ein Teil der Rohdaten. Auch undurchsichtige Copyright-Lizenzen sowie eingeschränkte Nutzungsbedingungen behindern einen freien Zugang und Umgang mit den öffentlichen Daten (vgl. Paderta 2012).

Es gibt jedoch bereits vorbildhafte Umsetzungen in den USA und England, die als Vorreiter anzusehen sind, wenn es um die Transparenz und die digitale Bereitstellung öffentlicher Daten im Zuge der Open-Government Bewegung geht. Tim Berners-Lee, Erfinder des World Wide Web, wurde im Frühjahr 2009 von Gordon Brown, dem damaligen Premierminister Großbritanniens gefragt, wie das Land die Möglichkeiten des World Wide Web besser nutzen könne. Seine Antwort darauf war, dass die Regierung alle ihre Daten offen zugänglich machen solle (vgl. Rogers 2010). Als Resultat lancierte im Januar 2010 das Datenportal *data.gov.uk*, welches Berners-Lee mitentwickelte, und Regierungsdaten online „und zwar nicht als PDF-Datei oder in einer Tabelle auf einer Website, sondern in einer Form, die eine einfache maschinelle Weiterverarbeitung ermöglicht“ (Schulzki-Haddouti 2010a), frei zugänglich machte. Vorbild für das Portal war die von der amerikanischen Bundesregierung gestartete Website *data.gov*, die schon im Mai 2009 ins Leben gerufen wurde. „Es war eine der ersten Amtshandlungen von US-Präsident Barack Obama, die Datenschätze der Regierung über den Datenkatalog ‚data.gov‘ frei zugänglich zu machen“ (Schulzki-Haddouti 2011). Seitdem haben Medien wie die Websites der *New York Times* und *The Guardian* die freigegebenen Daten der amerikanischen und britischen Regierungen auf den Portalen *data.gov* und *data.gov.uk* intensiv für ihre eigene datengestützte Berichterstattung genutzt.

3.1.2 Webscraping: Saugen aus dem Netz

„Da nicht alle Webseitenbetreiber offene Schnittstellen für ihre Inhalte anbieten, müssen Informationen extrahiert und in strukturierte Form gebracht werden“ (Fomin 2010: 17). Dabei kommt das sogenannte Screen-Scraping zum Einsatz. Screen-Scraping, zu Deutsch „Bildschirm auskratzen“, ist eine Methode zum automatischen Auslesen von Texten von Computerbildschirmen (vgl. Hoely 2011). In den meisten Fällen ist jedoch der Begriff Webscraping geläufiger, da das Screen-Scraping fast ausschließlich in Bezug auf Webseiten gesetzt und im Folgenden als Synonym benutzt wird. In diesem Rahmen heißt scrapen (schaben, kratzen) das Auffinden und „Einsammeln“ von Informationen von Websites, um diese in einem maschinenlesbaren Datensatz zusammenzuführen (vgl. Bergemann 2010); folglich das Kopieren von Daten aus dem Netz. Dabei werden die auf Websites in Textform gelagerten Informationen durch entsprechende Softwarewerkzeuge, Screen-Scraper oder Wrapper genannt, extrahiert und in einer Tabellen-

Datei gespeichert (vgl. Fomin 2010: 17). So können sie anschließend weiteranalysiert und verarbeitet werden.

Neben der Möglichkeit, selbst einen Wrapper zu programmieren, der einen vorher bestimmten Teil der Website identifiziert, kopiert sowie die Daten mittels eines Indexes anschließend untersucht und in einer Datenbank abspeichert, kann laut *Open Knowledge Foundation* (vgl. Lindenberg 2012) ein weniger technisch-versierter Datenjournalist kostenlose Tools wie Google Spreadsheets oder Yahoo Pipes benutzen, um Daten zu scrapen. Auch können Scraper-Plattformen wie zum Beispiel *scraperwiki.com* aufgesucht werden. Dort bieten Programmierer ihre Dienste für Datenextraktion kostenlos an und es stehen bereits fertig programmierte Wrapper kostenlos zur Verfügung.

Diese von Websites extrahierten Daten fallen in die zweite Schicht der in Kapitel 3.1 dargestellten Grafik der Datenzwiebel und werden als Transparent Data bezeichnet. Andreas Griebel erläutert, dass diese Transparent Data tendenziell auch „open“ sind, sich jedoch grundlegend in der Aufbereitung unterscheiden: „So liegt Transparent Data zwar frei zugänglich vor, jedoch nicht in Tabellen-Form und nicht in die gewünschte Relation gesetzt. Sie ist verstreut, nicht zusammengefasst, ihre Lizenzierung ist mitunter unklar“ (Griebel 2011). Als Beispiel dafür nennt er die Beziehungsstati auf Facebook. Sie sind laut Griebel prinzipiell zugänglich und anschaulich, aber sie liegen nicht als komplett vorhandener Datensatz vor, der über eine offene Schnittstelle heruntergeladen werden kann. So folgert er: „Du kannst sie erstellen, aber du müsstest sie erstellen“ (Griebel 2011).

Das Screen-Scraping ist für den DDJ eine exklusive Informationsquelle. Es bedeutet einerseits Unabhängigkeit von staatlicher Datenaufbereitung (vgl. Bergemann 2010), andererseits können Journalisten durch diese Technik selbst investigativ Daten aus dem Netz „ziehen“ und dadurch genügend Informationen sammeln, um eine umfassende und korrekte Auswertung der einzelnen Daten in einer datengestützten Berichterstattung umzusetzen. Der auf DDJ spezialisierte Glen McGregor beschreibt das Potential von Web-scraping folgendermaßen: „For data journalists, web scraping can be a powerful tool that allows them to assemble electronic records to find stories that could not be easily obtained otherwise“ (McGregor 2012).

Das Webscraping gewinnt dabei im wirtschaftlichen Sektor immer mehr an Bedeutung und führt zu einem rasch wachsenden Datenhandel. So sollen sich laut einem Online-Artikel aus dem *Wall Street Journal* (Angwin; Stecklow 2010) die Ausgaben für den Kauf von Daten aus dem Internet von 410 Millionen US-Dollar im Jahr 2009 auf 840 Millionen US-Dollar im Jahr 2012 verdoppeln. Dabei wird die Technik nicht immer zu rein ideellen und investigativ journalistischen Zwecken wie beim DDJ verwendet und steht deshalb in der Kritik. Einige Firmen verwenden das Scrapen dafür, gezielt geschützte, private Daten von Nutzern zu erheben wie z.B. aus Foren, Online-Datenbanken und sozialen Netzwerken (vgl. Trautmann 2011). „Das Kopieren von Daten im Internet ist nämlich nach wie vor eine juristische

Grauzone“ (Hoely 2011), da Urheberrechtsfragen für das automatische Auslesen von Daten auf Websites noch nicht eindeutig geklärt sind.

3.1.3 Leaked Information

Eine weitere Datenquelle für den Datenjournalismus stellen neben den in Kapitel 3.1.1 genauer beschriebenen Open Data und den aus dem Netz „gesaugten“ Transparent Data (siehe Kapitel 3.1.2) die sogenannten Leaked Data dar. Das Wort Leaks wird aus dem Englischen abgeleitet und bedeutet soviel wie „Lecks“, „Löcher“ oder „undichte Stellen“. Der Vorgang, der zur Leaked Data führt, wird als Data Leakage bezeichnet. Data Leakage bedeutet die nicht autorisierte Übertragung von organisationsinternen Daten an ein externes Ziel oder einen externen Empfänger (vgl. Gordon 2007: 5). Als Synonym für Data Leakage ist auch das Wort Information Leakage geläufig.

Im Zusammenhang mit dem hier dargestellten DDJ spielen dabei Enthüllungsplattformen, die sich im World Wide Web gebildet haben, wie zum Beispiel *Wikileaks* und *Open Leaks* eine zentrale Rolle für die datengestützte Berichterstattung, da sie riesige Dokumentensätze, die nicht zur Veröffentlichung vorgesehen waren, von teilweise anonymen Whistleblowern zentral auf einer Plattform für journalistische Zwecke zur Verfügung stellen. So beschreibt *Wikileaks* in einer aufgesetzten Definition über sich selbst:

We provide an innovative, secure and anonymous way for sources to leak information to our journalists (our electronic drop box). One of our most important activities is to publish original source material alongside our news stories so readers and historians alike can see evidence of the truth. (*Wikileaks* 2012).

Laut Andreas Grieb ist Leaked Data eine „Mischform aus Closed Data und Open Data. Genauer: Sie ist Closed Data, die Open wurde“ (Grieb 2011). Auch mit dieser Prämisse verfolgt man mit Leaked Data grundsätzlich das gleiche Ziel wie bei Open Data, nämlich Daten zu öffnen, sie jedem frei zugänglich zu machen und die Datensätzen dabei in strukturierter, maschinenlesbarer Form oder über eine offene Schnittstelle (API) zur Weiterverarbeitung und Weiterverbreitung anzubieten (vgl. Paderta 2012). Leaked Data dienen also in ihrem Grundsatz der Transparenz und Freiheit von digitalen Informationen. Sie sind jedoch auch Closed Data in dem Sinne, dass die Daten zwar für einen bestimmten Zweck erhoben wurden, sie jedoch unter Verschluss gehalten werden und ihre Publikation nicht für die Öffentlichkeit bestimmt ist. So entstehen bei der Verwendung von Leaked Data, genau so wie bei Transparent Data, einige Einschränkungen für den DDJ, besonders in Bezug auf ethische und rechtliche Fragen. Zudem sind ihre Vollständigkeit und ihr Kontext oft nicht sicher abzuklären. Laut Grieb ist eine Einordnung deshalb schwierig „und stellt sie immer unter Vorbehalt. Auch die Motivation des Whistleblowers und möglicher Zwischenkanäle färbt das Material“ (Grieb 2011). Eine weitere Hürde lässt sich am Beispiel der Ver-

öffentlichung der Afghanistan-Protokolle im Sommer 2010 von *Wikileaks* nachvollziehen: In diesem Fall waren die Daten nur begrenzt öffentlich, da die Datensätze zunächst nur speziellen Medienpartnern, dem *Spiegel*, *The Guardian* und der *New York Times* zugänglich gemacht wurden (vgl. Matzat 2011c). Ein solcher Zwischenfilter verstößt damit vom Prinzip her gegen die Auffassung des Datenjournalismus, Informationen jedem mit einer größtmöglichen Transparenz zugänglich zu machen (siehe Kapitel 2.3.1). Zudem hängt die Nutzung von Leaked Data stark von dem Veröffentlichungsformat ab, da Leaked Data prinzipiell nicht als Open Data gedacht sind, dadurch oft in nicht maschinenlesbaren Formaten vorliegen und somit per se nicht passend für eine datengestützte Berichterstattung sind (vgl. Gieß 2011). Der Umgang mit diesen Daten sollte also unter Vorbehalt stattfinden und intensiv von den Datenjournalisten geprüft und diskutiert werden. Datenjournalismus-Experten schließen Leaked Data als Datenquelle jedoch nicht aus, betonen aber, dass „die „reine Lehre“ des Datenjournalismus [...] ohnehin meist von öffentlich zugänglichen Daten wie etwa Open Government Data“ (Reiter 2012) spricht und Leaked Data deshalb eine größtmögliche Überprüfung erfordern. Leaked Data bieten jedoch im Kontext der in Kapitel 3.1 dargestellten Datenzweifel das größtmögliche Potential für den Datenjournalismus, investigativ zu recherchieren, zu arbeiten und damit eines der in Kapitel 2.3.1 erwähnten Charakteristika des Datenjournalismus zu erfüllen: „Die Daten stehen im Mittelpunkt einer investigativen Recherche“ (Langer 2011b: 5).

3.2 Datenanalyse und -aufbereitung

„Are you ready to be a journalist of the future? Better get your head wrapped around the idea of data“ (Arthur 2010), sagte der Begründer des World Wide Web Tim Berner-Lees 2010 in einem Artikel der Website der britischen Tageszeitung *The Guardian* und sieht Journalisten damit in der Pflicht, sich mit Daten und deren Analyse auseinanderzusetzen. „Sie hätten, so der Wissenschaftler, ‚datenhungrig‘ zu sein“ (Matzat 2012b). Dabei ist die Analyse der aggregierten Daten der Dreh- und Angelpunkt von datenjournalistischen Projekten. Mit der intensiven Auseinandersetzung von Daten entdecken Journalisten neues Wissen und Geschichten in ihnen, auch „auf die Gefahr hin, dass sie zwar viel Zeit investieren, aber nichts entdecken“ (vgl. Leßmöllmann 2012). Datenaufbereitung und -analyse ist somit einer der wichtigsten Schritte im DDJ, fordert jedoch von den Datenjournalisten gleichzeitig die intensivste Arbeit.

Welches Potential in der professionellen Analyse von Daten liegt, haben bereits viele internationale Medien und Verlagshäuser erkannt und drängen mit eigenen Abteilungen, die sich auf Datenjournalismus spezialisiert haben, auf den Markt. So unterhalten, wie bereits in Kapitel 1.1 beschrieben, der *The Guardian* und die *New York Times* eigene Ressorts – den sogenannten Datastore und Data-Desk – die sich um das Aggregieren, Gewichten und

Auswerten von Daten für redaktionelle Beiträge kümmern. Auch in Deutschland scheint sich der Trend DDJ zu etablieren, auch wenn er sich im Vergleich zu den angelsächsischen Medien weniger schnell entwickelt. „Zeit Online etwa baut seine Entwicklungsredaktion aus und versucht sich derzeit alle paar Wochen an einem datenjournalistischen Stück“ (Matzat 2012b), sowie die Berliner Tageszeitung *taz*, die einen eigenen Daten-Blog besitzt, auf dem Zahlen, Infos und Protokolle zu Reportagen und Analysen der Redakteure zum kostenlosen Download angeboten werden (vgl. *taz* 2012). Neben hauseigenen Abteilungen in Medien bieten auch externe Unternehmen ihre Dienste im DDJ an: Neben der *Deutschen Presse-Agentur*, die seit 2007 mit dem dpa-Data-Reporting eine eigene Abteilung betreibt, die sich allein mit datenjournalistischer Arbeit und deren Prozessen beschäftigt (vgl. dpa 2012), liefert die Datenjournalismus-Agentur *OpenDataCity*, unter der Feder der Journalisten Lorenz Matzat und Marco Maas, sowie dem Softwareentwickler Michael Kreil, fertig recherchierte und aufbereitete, datenjournalistische Stücke mit Visualisierungen und textlichen Elementen (vgl. Krampe 2012b).

Da es jedoch oft an der nötigen finanziellen Kapazität und personellen Ressourcen mangelt, kann sich nicht jede journalistische Redaktion eine eigene spezialisierte Abteilung oder das datenjournalistische Arbeitsstück eines externen Unternehmens leisten. „Es müssen aber nicht gleich ganze Datenredaktionen sein“ (Matzat 2012b), meint Lorenz Matzat. Derselben Meinung ist auch Ulrike Langer: „Es gibt immer mehr kostenfreie Software zur Datenfilterung und -präsentation“ (Langer 2011a: 3), die auch einzelnen Journalisten und kleinen Redaktionen erlaubt, professionellen Datenjournalismus, ohne großen finanziellen Aufwand zu betreiben. Auch müssen – da sind sich Experten wie zum Beispiel Ulrike Langer (vgl. Langer 2011c), Lorenz Matzat (vgl. Matzat 2011a) oder Paul Bradshaw (vgl. Bradshaw 2010) einig – die Datenjournalisten nicht als Programmierer fungieren. Sie sollten vielmehr „Daten beschaffen, Daten bewerten und den wesentlichen Punkt finden, der wirklich neue Einblicke, Relationen und Perspektiven bietet“, meint der Journalist und Informationsarchitekt Mirko Lorenz (vgl. Kaute 2012). Sie sollen „verstehen, ob ein geplantes Projekt einfach oder komplex zu programmieren ist“ (Kaute 2012), um zu kalkulieren, welches Budget für die Technik und den anstehenden Aufwand benötigt wird. Ein Beispiel für ein kostenloses Tool, welches gleichzeitig keine Programmierkenntnisse erfordert, ist der gemeinsam von Mirko Lorenz und dem französischen Programmierer Nicolas Kayser-Bril konzipierte Datawrapper: „Das Projekt ist Open-Source und in einer webbasierten und einer herunterladbaren Version frei verfügbar“ (Torial.com 2012). Ein weiteres Aufbereitungstool bietet der Suchmaschinengigant Google mit seinen Diensten Google Fusion Tables, Google Public Data Explorer und Google Text & Tabellen, die ermöglichen, kostenlos datenjournalistisch zu arbeiten (vgl. Holzinger 2012). Zudem stehen dem Datenjournalisten ausführliche und qualitativ hochwertige How-to-Anleitungen und weitere Empfehlungen für Tools auf

Websites, wie zum Beispiel auf der des *The Guardian* (vgl. Bradshaw 2010), dem *Datenjournal* (vgl. Holzinger 2012) oder im kürzlich erschienen „Data Journalism Handbook“ (vgl. Aisch 2012) zum Selbststudium zur Verfügung.

Trotz dieser hilfreichen Tools und frei verfügbaren Anleitungen sind statistisches Wissen und Tabellenkalkulation Grundvoraussetzungen für Datenjournalisten: „Man sollte wissen, wie man mit einer Tabelle umgeht und ein grundsätzliches Verständnis für Statistik mitbringen. Man sollte wissen, wie man Daten in eine Tabelle importiert und eine Datenbank wie Access bedienen können, mit der man einfache Suchabfragen oder Sortierungen vornimmt“, sagt Aron Pilhofer, leitender Datenjournalist der *New York Times* (vgl. Matzat 2012a). Datenjournalisten müssen deswegen im Umgang mit Tabellenkalkulationsprogrammen wie Excel, OpenOffice oder Numbers geübt sein. Zudem sollten Datenjournalisten nach Matzat (vgl. Matzat 2011a) die Fähigkeit der „computer-literacy“ beherrschen, folglich den Computer und seine Software sicher als Informationswerkzeug nutzen können, um dadurch Wissen zu generieren (vgl. Dewe; Weber 2007: 45). Auch kann das Wissen über Usability-Forschung im World Wide Web und über die Nutzung von interaktiven Infografiken für den Datenjournalisten bei seiner Arbeit sehr von Vorteil sein.

[...] Sie müssen wissen, wie man mit diesen Geschichten erzählt und was im Kopf eines Betrachters passieren (soll), der sich durch eine komplexe Grafik klickt [...]. Ohne den kompetenten Umgang mit multimedialen narrativen Strukturen wird der User im Wust bunter Kästchen untergehen. (Leßmöllmann 2012)

3.3 Erzählformen und Methoden

Die verschiedenen Erzählformen und Methoden des DDJ lassen sich laut Lorenz Matzat in sechs verschiedene Formen unterteilen, nämlich in Data Storytelling-, Echtzeitdaten-, Datensatz-, Crowdsourcing-, Hyperlokal- und Newsgames-Werke. In der folgenden Beschreibung dieser Formen wird nicht näher auf die Newsgames eingegangen, da es sich dabei um keine geschlossene Erzählform des Datenjournalismus handelt, sondern um das Prinzip „Gamification“ mit Nachrichten (vgl. Matzat 2011a). Eine genauere Erläuterung dieser Thematik würde den Rahmen der Arbeit übersteigen.

Angelehnt an Definitionen und Charakteristika des Datenjournalismus (siehe Kapitel 2.3 und 2.3.1) soll darauf verwiesen werden, dass es sich bei den kommenden Beschreibungen nur um eine Annäherung an die verschiedenen Erzählformen und Methoden handelt und nicht um verbindliche Muster, da oftmals Mischformen der nachfolgend beschriebenen Kategorien existieren und die Grenzen zwischen ihnen fließend verlaufen. So kann beispielsweise ein und derselbe Rohdatensatz für ein Data Storytelling-, Datensatz- und Hyperlokal-Werk verwendet werden.

3.3.1 Data Storytelling

Im Datenjournalismus kann manchmal der Eindruck erweckt werden, dass es allein um die ästhetische Präsentation der Daten geht und datenjournalistische Erzählformen nur dem Zweck nachgehen, Datenvisualisierungen oder interaktive, durchsuchbare Datenbanken zu produzieren. Der Informationsspezialist und leitende Redakteur bei den *BBC News*, Martin Rosenbaum beschreibt in seinem Artikel "Data Stories" im *Data Journalism Handbook*, dass die Methoden zwar sehr wertvoll für den DDJ seien und Aspekte eines sonst nicht überschaubaren Datensatzes prägnant und vor allem schnell für den Nutzer erfassbar machen, sie jedoch seiner Ansicht nach nicht den Kern bilden, der zu einer datengestützten Erzählform führt (vgl. Rosenbaum 2012). Es geht letzten Endes immer darum, einen Zusammenhang im Datensatz herzustellen, ergo die Geschichte in den Daten zu finden und sie zu erzählen: "All this can be very valuable, but like other forms of journalism, data journalism should also be about stories" (Rosenbaum 2012).

Laut Matzat besteht deshalb eine wesentliche Methode des DDJ aus dem geschichtenzentrierten Data Storytelling. „Hier meint Storytelling, dass eine Geschichte erzählt wird indem der Nutzer sich in dem gesetzten Rahmen nach eigenem Gusto über Details und Hintergründe eines Vorgangs erkundigen kann“ (Matzat 2011a). Der Rezipient muss nicht mehr – wie sonst bei einem starren Textbeitrag üblich – den linearen Erzählweg des Autors nachvollziehen, sondern erlebt den Beitrag (und die Geschichte um ihn herum) im Data Storytelling selbstständig im non-linearen Format. Er kann sich online durch die Datenbank oder –Visualisierung bewegen, „sich Ausschnitte betrachten oder Zusatzinformationen abrufen. Der Nutzer kann selbst entscheiden, wie tief er in die Materie einsteigen möchte“ (Matzat 2011a). Der Begriff Data Storytelling impliziert dabei bereits das Wort Storytelling und verfolgt gleichnamiges Prinzip: Er erzählt eine Geschichte, die den Zuhörer aktiv miteinbezieht (vgl. Suppan 2012). Durch die multimediale und interaktive Aufbereitung der Daten vollzieht der Nutzer so die Geschichte selbstständig nach und „bekommt die Informationen nicht nur vorgekaut“ (Macromedia Hochschule für Medien und Kommunikation 2012).

Ein treffendes Beispiel dafür, zeigt die Onlineausgabe der Wochenzeitung *Die Zeit* im Frühjahr 2011 am Beispiel der Berichterstattung „Verräterisches Handy“ (vgl. *zeit.de* 2011). Das Projekt soll die Auswirkungen des im Jahr 2007 verabschiedeten Gesetzes zur Vorratsdatenspeicherung, nach dem „Telefon- und Internetverbindungsdaten [...] ein halbes Jahr gespeichert werden“ (Focus Online 2012) dürfen, auf die Privatsphäre der deutschen Bevölkerung verdeutlichen. Im Zuge dieses Projekts stellte der Grünen-Netzpolitiker Malte Spitz nach erfolgreichem Einklagen von sechs Monaten seiner Ortsdaten bei der Telekom die gespeicherten Vorratsdaten *zeit.de* zur Verfügung. *zeit.de* verarbeitete zusammen mit dem Unternehmen *OpenDataCity* das Excelldokument mit 35.831 Zeilen (vgl. Biermann 2011) in einer interaktiven Datenvisualisierung „auf der die Nutzer alle Aufenthaltsorte, Telefonate und SMS des Politikers nachvollziehen [...]“ (Grimme

Online Award 2012) und ein halbes Jahr von Spitz' Leben, durch selbstständiges Navigieren innerhalb der interaktiven Karte, nach eigenem Belieben nacherleben konnten.

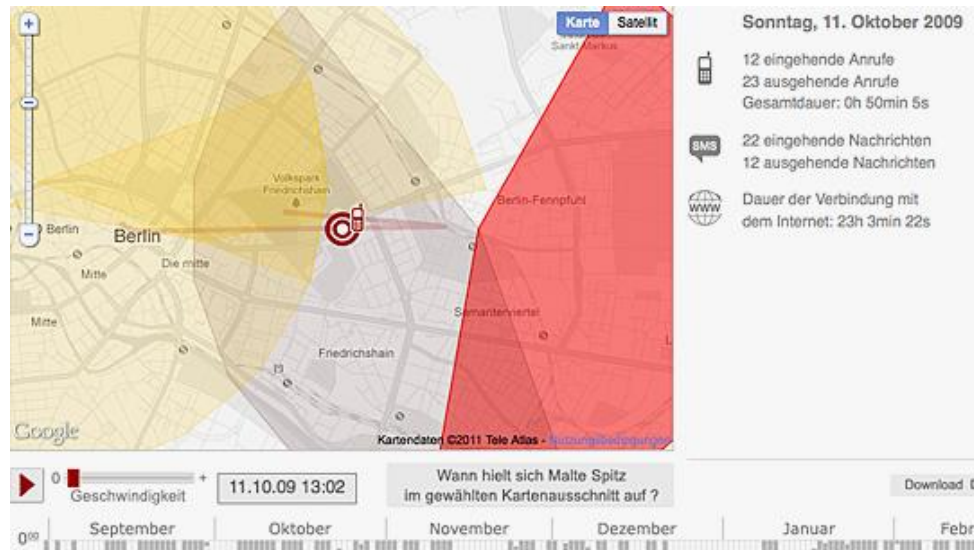


Abbildung 3: Visualisierung „Verräterisches Handy“ auf *zeit.de* auf Basis der Handydaten des Politikers Malte Spitz. Quelle: *zeit.de* 2011.

Angereichert wurde diese Visualisierung mit frei verfügbaren Inhalten aus dem Internet, wie Tweets und Blogbeiträgen von Spitz, Links zu Websites, auf denen der Politiker erwähnt wurde, Videos zum Thema, sowie dem ergänzenden Artikel „Was Vorratsdaten über uns verraten“ von *zeit.de*-Redakteur und Projektmitglied Kai Biermann (vgl. Langer 2011b: 5), der dort die Entwicklungsgeschichte der Datenvisualisierung und den thematischen, sowie rechtlichen Hintergrund der Daten erläutert. Auf die Grundlagen des Data Storytellings angewendet, ergeben sich folgende Punkte aus diesem Beispiel:

- Es wird eine Geschichte erzählt: Die Auswirkungen des 2007 verabschiedeten Gesetzes zur Vorratsdatenspeicherung und die Kritik am „gläsernen Bürger“.
- Ein Rahmen wird definiert: Über sechs Monate gespeicherte Vorratsdaten des Grünen-Politikers Spitz.
- Es erfolgt eine selbstständige Rezeption und das Nacherleben der datenvermittelten Geschichte: Mittels einer interaktiven Landkarte kann der Nutzer sechs Monate des Datenweges von Spitz nach eigenem Verständnis verfolgen.

- Der Nutzer kann selbst entscheiden, wie tief er in die Geschichte eindringen möchte: Es stehen verknüpfte multimediale Inhalte, wie Tweets und Videos als Hintergrundwissen zur Verfügung.

Durch diese Zusammensetzung aus datengestützter sowie interaktiver Visualisierung in Kombination mit klassischer Berichterstattung und multimedialen Inhalten veranschaulicht dieses Data-Storytelling-Werk die Problematik des Themas wesentlich eindringlicher, als es in reiner Textform durch Artikel hätte vermittelt werden können (vgl. Langer 2011b: 5). Die Rohdaten, die als Basis für die Datenvisualisierung verwendet wurden, stehen als Google-Spreadsheets im Sinne der Charakteristika des DDJ der Allgemeinheit frei zum Download zur weiteren Verwendung und Verarbeitung zur Verfügung (vgl. Biermann 2011). Für diesen Beitrag gewann das Projekt 2011 den Grimme Online Award in der Kategorie „Spezial“.

3.3.2 Echtzeitdaten

The notion that news comes in and stories go out — text and photos come in and paper goes out — is an artifact of the means of production and distribution, of course. Now a story never begins and it never ends. (Jarvis 2008)

Der US-amerikanische Journalist und Journalismus-Professor an der New Yorker Universität Jeff Jarvis beschreibt mit diesem Zitat eines der wichtigsten Merkmale von Echtzeitdatenwerken: Im Gegensatz zum Data Storytelling, bei dem die Berichterstattung ein fertiges Produkt darstellt, welches einen in sich geschlossenen Rahmen mit Anfang und Ende vorweist, entstehen Echtzeitdatenwerke im Prozess. Es ist eine „direkte, ungefilterte Berichterstattung vom Ort des Geschehens. Diese wird in kleinen Bröckchen ausgesendet und baut aufeinander auf“ (Schmidt 2011). Die Werke zeichnen sich durch den Faktor „Realtime“ aus und resultieren aus der Verwendung von Echtzeitdaten und deren Präsentation: [...] „The story evolves in real time rather than being presented as the full story at the end“ (Marks 2009). Auch nach Matzats Definition erkennt man Echtzeitdatenwerke daran, dass sie das gleichzeitige Sammeln von Echtzeitdaten und deren gleichzeitige Aufbereitung in einem Produkt ermöglichen (vgl. Matzat 2011a). Der Datenjournalismusexperte führt dazu ein Echtzeitdatenwerk der *The New York Times* aus dem Jahre 2010 an (vgl. *The New York Times* 2010). Die Tageszeitung aus New York City las während der Zwischenwahlen für das Repräsentantenhaus alle Kurznachrichten über die Wahl im sozialen Netzwerk Twitter aus, sammelte zehntausende Tweets von und an Kandidaten, sortierte sie und bereite diese in Echtzeit innerhalb einer Visualisierung auf (vgl. Matzat 2012: 4). „So erlaubte sie sowohl in Echtzeit – aber auch im Nachhinein – die Reaktionen im sozialen Netzwerk rund um Wahlkampf und Wahltag nachzuvollziehen“ (Matzat 2011a).

Auch Lorenz Matzat selbst war an der Entwicklung und Umsetzung eines Echtzeitdatenwerkes beteiligt. Zusammen mit seiner Datenjournalismus-Agentur OpenDataCity erstellte Matzat für den Online-Auftritt der *Süddeutschen Zeitung* das in seiner Form bislang einzigartige Projekt „Zugmonitor“ (vgl. Plöching 2012a).

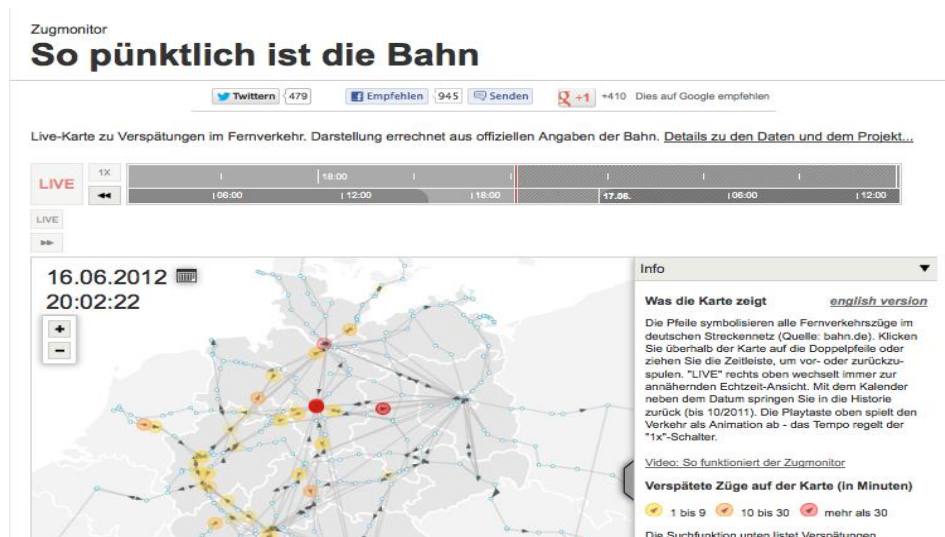


Abbildung 4: Das Projekt „Zugmonitor“ auf *zeit.de*. Quelle: *zeit.de* 2012.

Dort wird der Fernverkehr der Deutschen Bahn „anhand der Pünktlichkeitsangaben auf der Bahn-Website mitgeschnitten. In Echtzeit werden die Verspätungen der Züge kartiert“ (Matzat 2012: 4) und in einer ständig aktualisierten Grafik dargestellt. So entsteht ein für jeden zugängliches Live-Abbild des gesamten Schienennetzes der Deutschen Bahn, auf dem sich der Rezipient Verspätungsranglisten für Bahnhöfe und Züge anzeigen lassen, sich innerhalb des Archivs bewegen und sich über ein halbes Jahr Bahnverkehr im Zeitraffer nachvollziehen kann (vgl. Matzat 2012: 4).

3.3.3 Datensätze

Datenjournalismus-Werke enden nicht immer zwangsläufig in der Darstellung einer interaktiven Visualisierung. So beschreibt Matzat, dass das Bereitstellen von strukturierten Datenbanken ebenfalls als datenjournalistische Erzählform betrachtet werden kann (vgl. Matzat 2011a). Er betont jedoch, dass die zusammengetragenen Informationen durch eine journalistische Rechercheleistung erbracht und auf journalistische Sorgfaltspflicht hin geprüft worden sein müssen.

Diese Form äußert sich in sogenannten Datensatzwerken, die dem Prinzip des strukturierten Journalismus, mitunter als *structured journalism* bekannt, folgen: „Bei diesen Methoden werden Informationen in einem Sys-

tem organisiert und über eine Datenbankschnittstelle abrufbereit gemacht“ (Kayser-Brill 2011). Es entsteht eine Datenbank, die journalistisch verfasst, gepflegt und auf der Web- und Mobile-Anwendungen aufgebaut werden können (vgl. Chua 2011). Die Wochenzeitung *Die Zeit* stellte im Jahr 2010 ein solches Datensatzwerk auf ihrer Website online. Dabei handelte es sich um interaktive Datenbanken über die Spenden der Parteien von 2002 bis 2009, die aus der Kombination der vom Bundestag veröffentlichten und aus selbst aggregierten Daten von *zeit.de* aufbereitet wurden (vgl. Schulzki-Haddouti, 2010b).

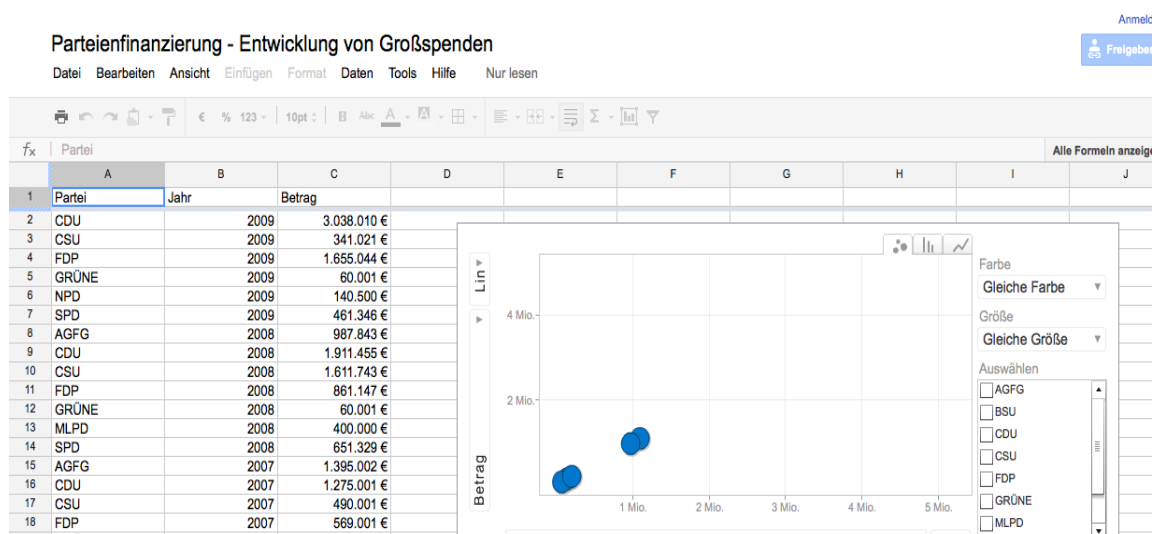


Abbildung 5: Interaktive Datenbank über die Parteispenden von 2002 bis 2009 auf *zeit.de*. Quelle: Schulzki-Haddouti, 2010b.

Mit den drei dort angebotenen Datenbanken gibt *zeit.de* den Nutzern die Möglichkeit, sich innerhalb einer interaktiven Rechercheumgebung neues Wissen anzueignen. Sie können sich die Daten selbst mit Hilfe der Google-Bewegungscharts erschließen, sie analysieren und kommentieren (vgl. Schulzki-Haddouti 2010c). Ein weiteres Beispiel wäre *The Guardian*, der auch im Bereich der Datensatzwerke seine Vorreiterrolle zeigt: Auf dem hauseigenen DataBlog werden nahezu jeden Tag ein oder mehrere Datensätze zu tagesaktuellen Geschehen zur Verfügungen gestellt (vgl. Schulzki-Haddouti 2012).

3.3.4 Crowdsourcing

„Crowdsourcing“ gilt als Hoffnung für investigativen Journalismus“ (Schuler 2009). Dabei beschreibt dieser Begriff eine Methode, auf die Journalisten bereits seit dem Ursprung dieses Berufszweigs zurückgreift: Sie bitten Informanten um Hinweise.

Der Begriff Crowdsourcing wurde 2006 durch den Journalisten Jeff Howe geprägt, der das neue Phänomen des Outsourcings an eine Menschenmenge auf der Internetseite des Magazins *Wired* beschrieb (vgl. Stanoevska-Slabeva 2011: 7). Howe definierte den Term dort als neue Quelle billiger Arbeitskräfte (vgl. Howe 2006b). In seinem Blog geht Howe etwas differenzierter auf die Definition des Terms ein: "Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call" (Howe 2006a). Crowdsourcing ist nach dieser Definition folglich eine klassische Form des Outsourcings. Es werden „auszulagernde Aufgaben per offenen Aufruf an eine große unbekannte Masse vergeben“ (Lang 2010: 3). Im World Wide Web verfolgt Crowdsourcing das Prinzip, dass eine undefinierte Zahl an Internetnutzern gemeinsam aktiv eine Aufgabe löst und die erworbenen Erkenntnisse an einem zentralen, öffentlich einsehbaren Ort dokumentiert.

Diese Methode wird im DDJ genutzt, wenn es um die Analyse großer Datenmengen geht, deren Analyse und Aufbereitung die Kapazität einer einzelnen Redaktion übersteigen würde. Simon Rogers von *The Guardian* beschreibt im *Data Journalism Handbook*, den Einsatz von Crowdsourcing im DDJ folgendermaßen: "Sometimes you will get a ton of files, statistics, or reports which it is impossible for one person to go through. Also you may get hold of material that is inaccessible or in a bad format and you aren't able to do much with it. This is where crowdsourcing can help" (Bouchart 2012). So entstehen die sogenannten Crowdsourcing-Werke, die nur durch und mit dessen Prinzip entstehen können. „Im großen Stil war es wieder der Guardian, der das Prinzip, das die Wikipedia groß gemacht hat, für den Journalismus einsetzte“ (Matzat 2011a), denn zum ersten Mal im investigativen Journalismus wurde durch die aktive Einbeziehung und Mitarbeit der Internetgemeinschaft eine umfangreiche Berichterstattung auf Basis eines Datensatzes ermöglicht (siehe Kapitel 2.3.2).

Im Frühjahr 2009 machte die britische Tageszeitung von sich reden, als sie 458.000 Dokumente zum britischen Spesenskandal der Unterhausabgeordneten zum Herunterladen ins World Wide Web stellte. Die Dokumente wurden zuvor vom Parlament veröffentlicht. Sie lagen jedoch teilweise in handschriftlicher Form vor und waren somit maschinell nicht auswertbar (vgl. Matzat 2011a). Zudem fehlte es *The Guardian* an personellen Ressourcen, um den unstrukturierten Datensatz in Eigenarbeit zu analysieren. Er bat deshalb ihre Nutzer in ihrem Projekt "Investigate your MP's Expenses": „Können Sie uns mit der Überprüfung dieser Daten helfen?“ (vgl. Lisakowski 2012).

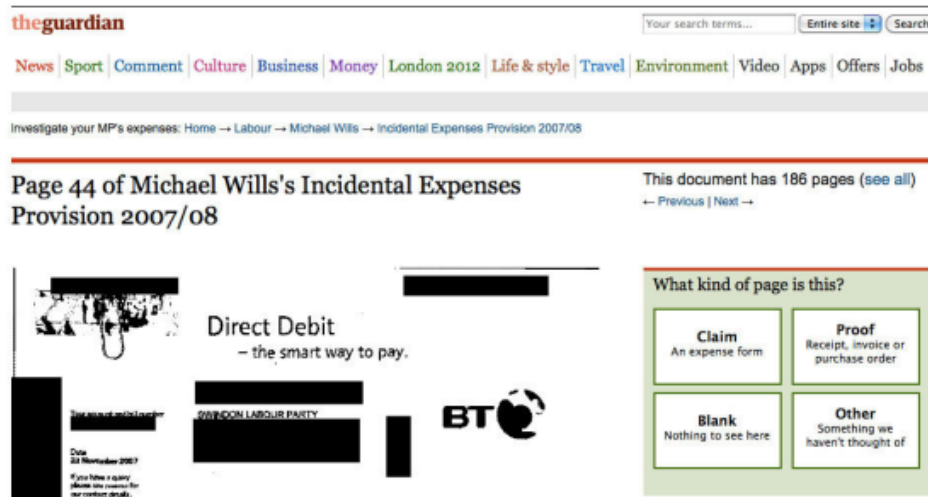


Abbildung 6: Das *Guardian*-Projekt "Investigate your MP's Expenses" zum Spesenskandal in Großbritannien. Quelle: guardian.co.uk 2012.

Um den Nutzer besser in die Medienanalyse einzubeziehen und ihm die Arbeit zu erleichtern, erstellte die Tageszeitung eine eigens dafür konzipierte Webapplikation zur Handhabung und Analyse der Datensätze. Die Reaktionen auf dieses Projekt waren laut dem mitwirkenden Journalisten Simon Rogers überwältigend (vgl. Bouchart 2012). Über 28.000 Internetnutzer halfen online mit und werteten fast 50 Prozent der Spesenskandal-Dokumente aus (vgl. guardian.co.uk 2009), um die Redaktion auf mögliche Verfehlungen der Politiker hinzuweisen (vgl. Matzat 2011a). Die daraus gewonnenen Erkenntnisse publizierte *The Guardian* sowohl im Print als auch auf seiner Website und deckte dadurch einen der größten Politikskandale Großbritanniens auf, der dafür sorgte, dass in Großbritannien noch im selben Jahr Rücktritte von Politikern und Gesetzesänderungen folgten (vgl. Oswald 2011).

In Deutschland führte ein solches datenjournalistisches Projekt, das „GuttenPlag Wiki“ dazu, dass dem deutschen Politiker Karl-Theodor zu Guttenberg der Dokortitel aberkannt wurde und er daraufhin als Verteidigungsminister zurücktrat. Bei dem Projekt ging es darum, die Dissertation zu Guttenbergs nach Plagiaten zu durchforsten und diese exakt zu dokumentieren. Auslöser für die Geburt von „GuttenPlag Wiki“ war ein am 17. Februar 2011 veröffentlichter Artikel der Süddeutschen Zeitung in dem Rechtswissenschaftler Andreas Fischer-Lescano, „einige grobe Fehler aufgefallen waren, nachdem er sich kritisch mit zu Guttenbergs Doktorarbeit auseinandergesetzt hatte“ (Klinger 2012). Nur einen Tag später startete das Projekt „GuttenPlag Wiki“ mit Hilfe des österreichischen Experten für Wissen-

schaftsplagiate (vgl. Neuroth 2012). Über Tausend Plagiatsjäger sichteten die 393 Seiten und fanden 94,4 Prozent plagiierte Passagen in der Arbeit (vgl. Neuroth 2012). Die Darstellung der kollaborativen Arbeit fand jedoch nicht nur auf reiner Textebene statt, sondern wurde auch visuell umgesetzt. Auf der Website „Interaktiver Guttenberg Report“ ist jede Seite je nach Plagiatsgrad weiß, hellrot oder dunkelrot eingefärbt (vgl. Oswald 2011). Dadurch wird dem Nutzer und den Helfern visuell verdeutlicht, welche Seiten Plagiate enthalten und wie viel Prozent die kopierten Textstellen ausmachen (siehe Abb. 7). „Die gefundenen Plagiatsstellen werden dem jeweiligen Original gegenübergestellt“ (Patzwald 2012). Genauso wie das in Kapitel 3.3.1 erläuterte datenjournalistische Werk „Verräterisches Handy“ von *zeit.de*, erhielt auch das „GuttenPlag Wiki“ im Jahr 2011 den Grimme Online Award.

Besonders dieses Beispiel zeigt, welches Potential in der Methode des Crowdsourcings für den Datenjournalismus liegt: Während das renommierte, deutsche Magazin *Der Spiegel* nach fünf Tagen erst 60 Plagiatsstellen aufdeckte, hatten die Plagiatsjäger auf „GuttenPlag Wiki“ weit mehr als 300 betroffene Textstellen gefunden (vgl. Pohlmann 2011). Crowdsourcing-Werke können damit Redaktionen helfen, ihre Leistungsgrenzen in der Datenanalyse zu überwinden. So betont Christian Stöcker, Ressortleiter Netzwelt von *spiegel.de*, dass „ohne die akribische Dokumentation der Plagiate im GuttenPlag Wiki [...] die Debatte versandet“ (Stöcker 2011) wäre und im öffentlichen Diskurs wohl eine andere Wende genommen hätte. So bieten datengestützte Crowdsourcing-Werke nicht nur Journalisten den Vorteil, schneller an relevante Informationen zu gelangen und die Recherche-Ergebnisse für ihre eigene Berichterstattung zu nutzen, sondern ermöglichen es zudem, Nutzer selbst in die Medienberichterstattung miteinzubeziehen, sie aktiv mitarbeiten zu lassen und sie dadurch für Datenarbeit zu sensibilisieren (vgl. Ruppert 2012).

Interaktiver Gutenberg Report

Eine graphische Aufbereitung der im GutenPlag Wiki gesammelten Daten zu Karl-Theodor Freiherr zu Guttenbergs Dissertation „Verfassung und Verfassungsvertrag“

vielen Dank an fiesh, vis4net, pettar, monopteros, Adalbert, Egon, kahrl und viele andere Wikianer!

Bei den folgenden Seiten handelt es sich um vollständig automatisch generierte Darstellungen der im [GutenPlag Wiki](#) gefundenen Textstellen.
Die farblich gekennzeichneten Seiten enthalten nicht ausgewiesene Zitate oder Plagiate aus anderen Veröffentlichungen.
Die dunkelroten Symbole kennzeichnen Seiten, auf denen Plagiate unterschiedlicher Quellen gefunden wurden.

(Die farbliche Hervorhebung unten basiert auf den Daten der im Wiki gesammelten Seiten. Diese sind noch nicht vollständig in das neue „Fragments“-Format übertragen, so dass auf manchen rot hinterlegten Seiten die graphische Darstellung von Plagiaten fehlen kann.)

Der interaktive Gutenberg Report darf als ausgelagerter Teil des Wikis verstanden werden. Sämtliche Inhalte dürfen daher unter Quellenangabe zu Presse Zwecken verwendet werden.

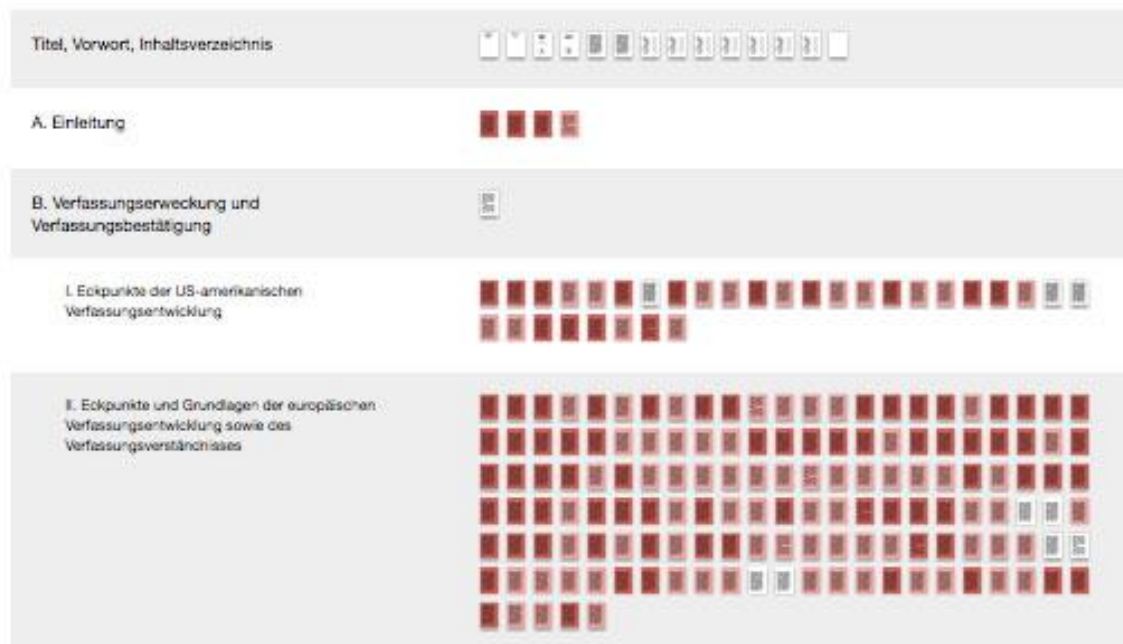


Abbildung 7: Grafische Aufbereitung der im „GutenPlag Wiki“ gesammelten Daten zu Karl-Theodor Freiherr zu Guttenbergs Dissertation „Verfassung und Verfassungsvertrag“. Quelle: GutenPlag 2012.

3.3.5 Hyperlokal

„Die klassische Lokalzeitung ist vom Aussterben bedroht. In den USA gibt es in vielen Landesteilen längst keine Regionalblätter mehr. Auch hierzulande haben etliche Verleger mit sinkenden Auflagen und einer alternden Leserschaft zu kämpfen“ (Langer 2012). Ein möglicher Weg aus der Krise könnten Hyperlokal-Werke sein, die sich auf Daten aus der unmittelbaren Umgebung beziehen (vgl. Matzat 2011d).

Den Rezipienten von lokaljournalistischen Angeboten interessiert, was sich in seiner unmittelbaren Umgebung abspielt und in seinem sozialen Umfeld thematisiert wird: „Der Verein um die Ecke, das Fest im nahen Kindergarten, die Baustelle vorne an der Kreuzung, der Einbruch beim Nachbarn gegenüber und das Angebot vom Laden die Straße herunter“ (Matzat 2011d). Mit den Möglichkeiten des Internets können solche Informationen kleinteiliger und spezifischer dargestellt werden: Nachrichten aus Politik, Medien, Kultur, Konsum und der Verwaltung können nicht bloß aus der eigenen Stadt, sondern aus dem eigenen Viertel, dem eigenen Straßenblock gefunden werden (vgl. Langer 2012). „Dazu lokale Meldungen aus sozialen Netzwerken oder von kommunalen Akteuren wie Behörden oder Einzelhändlern“ (Langer 2012), die auf einer einzigen Plattform vereint zu finden seien. Diesen Ansatz – das Darstellen von auf die Mikroebene von Straßen oder Häuserblocks bezogenen Daten – verfolgen Hyperlokale-Werke. Über mobile Internetzugänge kann der Nutzer zudem unmittelbar Informationen zu seinem Standort abrufen. In den meisten Fällen wird für die Darstellung eine interaktive Karte gewählt. Neben dem Begriff „Hyperlokal“-Werke wird auch der Terminus „Mikrolokal“-Werke benutzt (vgl. Matzat 2011d).

Pionierarbeit leistete dabei Adrian Holovaty, der bereits im Kapitel 2.3.2 im Kontext der Meilensteine des Datenjournalismus erwähnt wurde. Der Journalist und Programmierer Holovaty startete 2007 die Plattform *Everyblock*, die als öffentliche Datenbank für hyperlokale Informationen fungiert und mittlerweile in sechzehn großen Städten von New York über Chicago und Los Angeles bis nach Seattle erfolgreich läuft. Die Plattform produziert keine eigenen Inhalte, sondern sammelt sie von unterschiedlichen Quellen. „Man versteht sich als ‚geographischer Filter‘ heißt es in der Selbstdarstellung“ (Matzat 2010e).

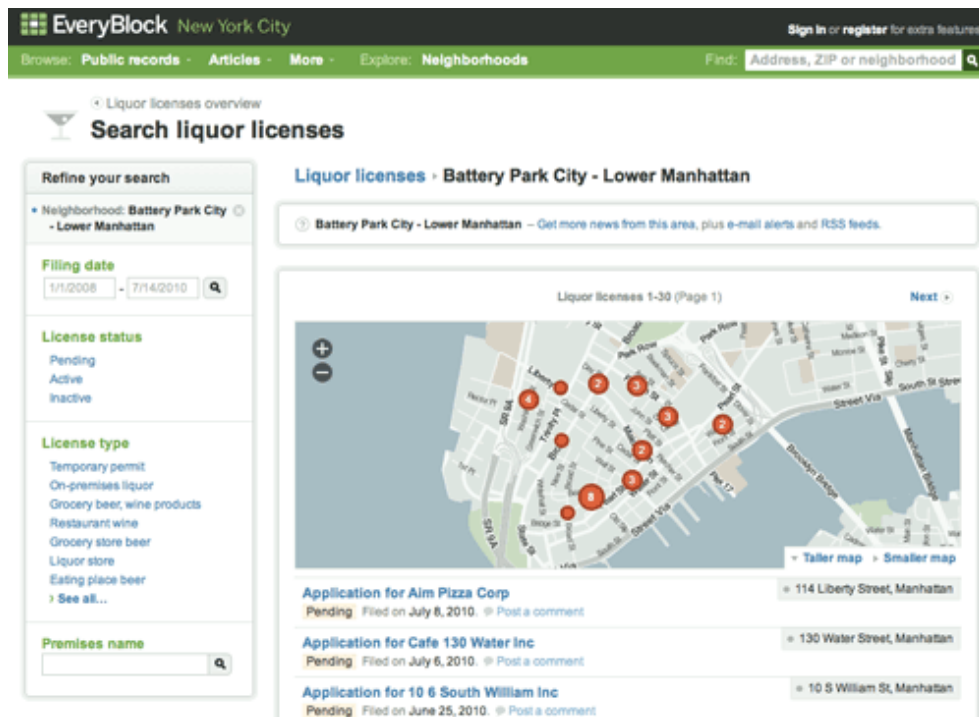


Abbildung 8: Vorreiterrolle im Bereich Hyperlokale-Werke: das Portal *Everyblock*. Quelle: Everyblock 2012.

Everyblock bezieht seine verwendeten Daten aus Ereignissen innerhalb der unmittelbaren Nachbarschaft. Laut Matzat geht es dabei jedoch nicht um statische Daten, wie Adressen von Schulen. „Sondern um Nachrichten im weitesten Sinne, die sich auf einen Standort beziehen und ein Datum haben“ (Matzat 2010e). In *Everyblock* werden unter anderem Nachrichten aus den Medien, politische Neuigkeiten, Graffiti-Entfernungen, Baugenehmigungen und Polizeiticker verwendet und Quellen wie Fotos von Flickr, Anzeigen von Craigslist, Blogartikel und OpenData hinzugezogen (vgl. Langer 2012).

Ein aktuelles Beispiel bietet das Hyperlokale-Werk „Frankfurt Gestalten-Bürger machen Stadt“. Hier kann sich der Nutzer straßengenau über die Lokalpolitik von Frankfurt informieren. Seit 2006 bereitet *Frankfurt-gestalten.de* die Informationen der Parlos-Datenbank der Stadt Frankfurt auf, verschlagwortet sowie georeferenziert die Daten und zeigt das Ergebnis anschaulich auf einer Karte (vgl. Kreutz 2010). Das Projekt soll dazu dienen, Bürger und Bürgerinnen zu vernetzen, um ihre Stadt zu gestalten (vgl. Frankfurt Gestalten 2012). So können die Nutzer sich über aktuelle Anträge der jeweiligen Stadtbeiräte informieren, diese diskutieren und auch eigene Initiativen starten. Neben dem Ziel, die Potenziale von offenen Daten aufzuzeigen, soll das Projekt zudem die politische Diskussionskultur der Bürger aufwerten und weiterführen.

4 DAS POTENTIAL DES DATENJOURNALISMUS

Im Zeitalter einer voranschreitenden Digitalisierung und Technisierung von Informationen steigt der Datenberg im World Wide Web kontinuierlich: Daten avancieren zum Rohstoff, mit dem sich viel Profit erzielen lässt (vgl. Palmer 2006). Lorenz Matzat erläutert diesen Zustand im Rahmen eines Vortrags auf der *localwebconference* 2012 folgendermaßen: „Es wird immer mehr Daten geben, weil es immer mehr Geräte gibt, die Daten sammeln und es gibt riesige Datenmengen in Unternehmen, Verwaltungen und durch Open Data aber auch im privaten Bereich. Kaum jemand kann sie noch wirklich durchdringen“ (Lokalrundfunktag 2012: 1:25 bis 2:10 Minute). Die digitale Revolution als zentrales Element unserer heutigen Gesellschaft ist somit gleichzeitig ein Kampf um die Informationshierarchie im Netz: „Wer kann Informationen zurückhalten, wer hat die Deutungshoheit über das, was geschieht. Und wer kann seine Deutung mit größtmöglicher und weitreichender Autorität verbreiten“ (Matzat 2010a). Lorenz meint, der Journalismus darf sich nicht vor der Auseinandersetzung mit Datenanalyse und -auswertung scheuen und muss auf die Entwicklungen der Datenzeit reagieren. Die Antwort auf diese neuen Herausforderungen, die an den Journalismus von morgen gestellt werden, ist der Datenjournalismus und das in ihm liegende Potential.

Die im Verlauf der durchgeführten Untersuchung gewonnenen Einsichten zeigen, dass die immer komplexer gewordene Gesellschaft anhand von DJJ präziser, schneller und umfangreicher dargestellt werden kann. „Datenjournalismus antwortet auf eine verdatete Welt, die ihre gesellschaftlichen Prozesse häufig in Datensätzen dokumentiert und ablegt – ein Fundus für relevante journalistische Geschichten“ (Leßmöllmann 2012). Eine datenjournalistische Berichterstattung kann somit neue Zusammenhänge und Fakten aufdecken, die sonst verborgen geblieben wären. Im Weiteren wird näher auf die einzelnen Punkte eingegangen, die das Potential, welches der Datenjournalismus im World Wide Web birgt, verdeutlichen.

4.1 Exklusivität und Unabhängigkeit

Nach Lorenz Matzats Ansicht schöpft der Online-Journalismus bislang noch nicht sein volles Potential aus. „Die Website einer Zeitung imitiert irgendwie immer noch das Zeitungsblatt. Ich denke Daten bieten eine neue Chance in diesem Bereich, weil sie dem Trägermedium Internet entsprechen“ erläutert Matzat. „Das Internet kann etwas, was all die anderen Nebenformen nicht können: journalistische Geschichten non-linear und interaktiv erzählen“ (Lokalrundfunktag 2012: 2:10 bis 2:50 Minute). Im DDJ liegt also die Kraft, das volle interaktive und multimediale Potential des Mediums Web zu nutzen. Journalistische Webpräsenzen könnten so gegenüber ihren gedruckten Mutterblättern ein Alleinstellungsmerkmal etablieren und ihre

Stärke ausspielen (vgl. Buettner 2012), da sie eine völlig neue Form der Berichterstattung ermöglichen (siehe Kapitel 2.3). Auch wollen Rezipienten von journalistischen Angeboten einzigartige Inhalte und keine Nachrichten, die sich überall im Netz wieder finden lassen, meint Aron Pilhofer (Langer 2011b: 7), Leiter der 14-köpfigen Abteilung Interactive News in der Redaktion der *New York Times*. Datenjournalistische Werke bieten dabei den Vorteil, einzigartig und in ihrer Form nicht leicht imitierbar zu sein: „Interaktive Datenanwendungen kann man nicht einfach kopieren. Ich kann nicht, wie ein Artikel oder Foto das einfach in den Zwischenspeicher meines Rechners legen und irgendwo anders einfügen, das ist nicht möglich“ (Lokalrundfunktag 2012: 18:00 bis 18:40 Minute). Ein journalistisches Medium, welches ein datenjournalistisches Werk publiziert, kann sich somit der Exklusivität seiner Geschichte sicher sein. Zudem sind datengestützte Berichterstattungen nachhaltig und aktualisierbar, da die Anwendungen auf Datenbanken basieren und auch nach deren Publikation mit weiteren Daten angereichert werden können. „Wenn im nächsten Jahr der Datensatz vom Vorjahr veröffentlicht wird, kann ich die Anwendung einfach damit ergänzen, die Anwendung ist somit nicht tot, sie sie kann je nach Thema immer weiter wachsen und zugänglich bleiben“ (Lokalrundfunktag 2012: 18:40 bis 19:00 Minute). Dadurch schafft der Datenjournalismus in seiner Form und Ausprägung Exklusivität im Bereich des Journalismus und Mehrwert für den Online-Journalismus.

DDJ schafft weiterhin Unabhängigkeit in der Berichterstattung, da sich Journalisten nicht mehr ausschließlich auf die Datenauswertungen von externen Unternehmen, Behörden oder aus der Wissenschaft verlassen müssen. Datenjournalisten können laut Christina Elmer (vgl. ZAPP 2011: 1:41 bis 2:00) durch die Fähigkeit Rohdatensätze zu analysieren und aufzubereiten, selbst entscheiden, was den Kern in eines Datensatzes ausmache und dadurch exklusive journalistische Geschichten schaffen. Dadurch findet man jedoch nicht nur exklusive Geschichten, sondern ermöglicht auch unabhängiges Recherchieren der Journalisten. „Wenn man Rohdaten selber hat, ist man unabhängig von denen, die die Daten auswerten“ (Focke 2011). So kann der DDJ Zusammenhänge aufzeigen, die vorher nicht ersichtlich waren und in dieser Form nicht hätten erzählt werden können. Der Journalist ist nicht mehr auf Statistiken und Diagramme aus zweiter oder dritter Hand angewiesen und kann alle Daten, die er benötigt, einfach selbst auswerten. „Was dem Journalismus in seinem Selbstverständnis als unabhängiger Vermittler durchaus entgegen kommen würde“ (Macromedia Hochschule für Medien und Kommunikation 2012).

Diese Exklusivität und Unabhängigkeit, die durch solche datenjournalistischen Arbeiten bewirkt wird bleibt, wie in Kapitel 3.2 beschrieben, nicht nur großen Medien- und Verlagshäusern vorbehalten. Durch das große Angebot an kostenfreien Software-Tools und frei zugänglichen Anleitungen können auch kleine Redaktionen und einzelne Journalisten DDJ betreiben.

4.2 Transparenz und Glaubwürdigkeit

Datenjournalismus verfolgt, wie in der Definition des Datenjournalismus beschrieben, den Grundsatz von Open Data und richtet sich hin zu einer freien und offenen Netzkultur: „Freier Zugang, das Teilen von Informationen und Zusammenarbeit der Internetnutzer“ (Matzat 2010a). So gehört zu der konsequenten Umsetzung von datenjournalistischen Projekten auch immer das Bereitstellen der dazugehörigen Rohdaten für den Nutzer (vgl. Langer 2011b: 4), folglich eine vollkommene Offenlegung der Daten. Der Datenjournalismus übt somit eine bisher in dieser Form noch nicht existente Quellen-Offenheit im Journalismus aus, die erlaubt, dass Rezipienten die Fakten einer Geschichte selbst prüfen und nachvollziehen können. Lorenz Matzat erläutert dazu, dass Datenjournalismus ein Format sei, welches der heutigen, selbstbestimmten Mediennutzung der Internetgesellschaft entgegen komme. DDJ reagiere auf Nutzer, die selbst Themen erkunden wollen und sich in einer möglichst kurzen Zeit, schnell und aus vielen Quellen Informationen beschaffen würden. „Für die ist ein normaler Text-Artikel nicht mehr das journalistische Format, welches sie konsumieren. Sie wollen selber aktiv werden und sich selbst mit der Thematik auseinandersetzen. Für solche Menschen ist eine interaktive Datenanwendung, zumindest aus journalistischer Sicht, sicherlich der richtige Weg“ (Lokalrundfunktag 2012: 19:00 bis 19:25). Dabei bewirkt die non-lineare Erzählform der DDJ-Werke, dass der Rezipient den Beitrag selbst erlebt und selbstständig erkunden kann: Dem Nutzer wird die Wahl gelassen, welche Ausschnitte eines datengestützten Werkes er rezipiert, wie tief er in die Materie einsteigt und welche Zusatzinformationen er abrufen kann. Der DDJ schafft dadurch Transparenz und baut gleichzeitig Vertrauen beim Rezipienten in die journalistische Berichterstattung und den Online-Journalismus auf. Diese Komponenten sind besonders wichtig im Hinblick auf das Zusammenspiel der Datenjournalisten mit der Internetgemeinde bei Crowdsourcing-Werken. Nur wenn die Masse der Internetnutzer Vertrauen in das Medium hat, wird sie zu einer kollaborativen Recherche und Analyse der Daten bereit sein und den Journalisten dabei helfen, schneller an relevante Informationen zu gelangen, welche als Recherche-Ergebnisse für datengestützte Berichterstattung genutzt werden.

Neben der Vertrauensbasis, die der DDJ zwischen Nutzern und Journalisten schafft, stärkt der DDJ auch die Glaubwürdigkeit in die Medien und den Journalismus. Durch die selbstständige Recherche der Daten und deren intensive Aufbereitung besinnt sich der DDJ auf den Ursprung des Qualitätsjournalismus zurück: Er konzentriert sich auf die Hintergrundberichterstattung, recherchiert sorgfältig, prüft Fakten und bereitet die Inhalte redaktionell-reflektiert auf (vgl. Thurner 2010). Web-Angebote mit datenjournalistischen Stücken kräftigen dadurch ihre eigene Rolle als Primärinformationsquelle sowie Kontrollorgan der Gesellschaft und können ihrem Angebot zu mehr Seriosität helfen. In Zeiten der „Gratiskultur“, in der Nutzer kostenlo-

sen journalistischen Inhalt gewohnt sind und journalistische Online-Angebote sich von News-Portalen, Blogs, PR-Seiten oder Bürgerjournalismus abzugrenzen versuchen, könnte die durch die hohe, qualitative Berichterstattung des Datenjournalismus gewonnene Glaubwürdigkeit helfen, journalistische Webangebote finanziell ertragreich zu machen (vgl. Gehmlich 2012).

5 FAZIT UND AUSBLICK

„Datenjournalismus ist gekommen, um zu bleiben“ (OpenDataCity 2012).

Zusammenfassend bleibt festzuhalten, dass der DDJ ein immenses Potential für den Journalismus und vor allem für den Online-Journalismus in sich trägt: Er bewirkt **Exklusivität**, ermöglicht **Unabhängigkeit**, schafft **Transparenz** und fördert **Glaubwürdigkeit**. Datenjournalismus schafft es, durch den professionellen Umgang mit Daten den Online-Journalismus zu verbessern und bietet die Möglichkeit, dass sich Journalisten an die aktuellen Entwicklungen des digitalen Zeitalters anpassen. Dabei kann Datenjournalismus nicht nur als vorübergehender Trend angesehen werden, sondern nimmt die Position einer neuen, journalistischen Avantgarde im World Wide Web ein, deren feste Etablierung in journalistischen Online-Angeboten in den nächsten Jahren zunehmend zu betrachten sein wird: Denn die Datenberge im World Wide Web werden nicht kleiner. Sie sind der Rohstoff, der den Journalismus beflügeln und ihm helfen wird, politische, soziale und wirtschaftliche Vorgänge auf eine neue, qualitativ-hochwertige und spektakuläre Weise abzubilden. Journalisten, die mit Daten umgehen können, sind heute schon gefragte Experten (vgl. Langer 2011b: 1).

6 LITERATUR- UND QUELLENVERZEICHNIS

Aisch, Gregor; Boyer, Brian; Doig, Steve; Evans, Lisa; Klein, Scott; Markun, Pedro; O'Murchu, Cynthia; Peralta Ramos, Angélica; Phillips, Cheryl (2012): Data Journalists Discuss Their Tools of Choice.

http://datajournalismhandbook.org/1.0/en/understanding_data_6.html [letzter Zugriff: 19.06.2012].

Angwin, Julian; Stecklow, Steve (2010): „Scrapers“ Dig Deep for Data on Web. <http://online.wsj.com/article/SB10001424052748703358504575544381288117888.html> [letzter Zugriff: 01.06.2012].

Ausserhofer, Julian (2010): Datenjournalismus: Meilensteine und Charakteristika.

<http://www.slideshare.net/julianausserhofer/datenjournalismus-meilensteine-und-charakteristika-11021907> [letzter Zugriff: 09.05.2012].

Arthur, Charles (2010): Journalists of the future need data skills, says Berners-Lee.

<http://www.guardian.co.uk/technology/organgrinder/2010/nov/19/berners-lee-journalism-data> [letzter Zugriff: 19.06.2012].

Beckedahl, Markus (2010): Momentum für Datenjournalismus: Die Irak-Protokolle.

<http://netzpolitik.org/2010/momentum-fur-datenjournalismus-die-irak-protokolle/> [letzter Zugriff: 20.05.2012].

Beckedahl, Markus (2012): Data Journalism Handbook.

<http://netzpolitik.org/2012/data-journalism-handbook/> [letzter Zugriff: 20.05.2012].

Bensch, Danielle (2011a): 2002 begann das digitale Zeitalter.

http://www.welt.de/print/die_welt/wissen/article12506319/2002-begann-das-Digitalzeitalter.html [letzter Zugriff: 28.04.2012].

Bensch, Danielle (2011b): Die Menschheit erstickt an ihren Daten.

<http://www.welt.de/wirtschaft/webwelt/article12507803/Die-Menschheit-erstickt-an-ihren-Daten.html> [letzter Zugriff: 02.05.2012].

Bergemann, Benjamin (2010): Datenjournalismus: Die zwei Seiten der Transparenz.

<http://politik-digital.de/datenjournalismus-die-zwei-seiten-der-transparenz/> [letzter Zugriff: 01.06.2012].

- Betterplace Lab (2012): Daten für Alle.
<http://trendreport.betterplace-lab.org/trend/daten-fuer-alle/> [letzter Zugriff: 29.05.2012].
- Biermann, Kai (2011): Was Vorratsdaten über uns verraten. In: Zeit online.
 24.02.2011.
<http://www.zeit.de/digital/datenschutz/2011-02/vorratsdaten-malte-spitz> [letzter Zugriff: 10.06.2012].
- Bouchart, Marianne (2012): Crowdsourcing Data at the Guardian Datablog.
http://datajournalismhandbook.org/1.0/en/getting_data_5.html [letzter Zugriff: 18.06.2012].
- Bouchon, Catherine (2006): Einsatz, Gestaltungsqualität und Informationsvermittlung von Infografiken. Bonn: GRIN Verlag.
- Bradshaw, Paul (2010): How to be a data journalist. In: The Guardian Datablog.
 01.10.2010.
<http://www.guardian.co.uk/news/datablog/2010/oct/01/data-journalism-how-to-guide> [letzter Zugriff: 09.05.2012].
- Bradshaw, Paul; Rohumaa, Liisa (2011): The Online Journalism Handbook. Skills To Survive And Thrive In The Digital Age. Harlow: Longman.
- Buettner, Mathias (2012): Geschichten aus 1001 Database. Daten sind Storys!
<http://blog.der13stock.de/2012/05/geschichten-aus-1001-database/> [letzter Zugriff: Abruf: 21.06.2012].
- Bundeszentrale für politische Bildung (2012): Dossier Open Source.
<http://www.bpb.de/gesellschaft/medien/opensource/> [letzter Zugriff: 29.05.2012].
- Cisco (2011): Cisco Visual Networking Index: Forecast and Methodology, 2010-2015.
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html [letzter Zugriff: 02.05.2012].
- Cisco (2012): Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2011-2016.
http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.html [letzter Zugriff: 02.05.2012].
- Chua, Reg (2011): Structured Journalism.
<http://structureofnews.wordpress.com/structured-journalism/> [letzter Zugriff: 16.06.2012].

Davis, Chares H.; Shaw, Debora (2011): Introduction to information science and technology.

<http://bit.ly/KphX4e> [letzter Zugriff: 02.05.2012].

Dean, Talisa (2012): Datenjournal. Online-Journalismus zwischen Daten und Multimedia.

<http://www.datenjournal.de/dossier/einstieg> [letzter Zugriff: 07.05.2012].

Definition of Free Cultural Works (2012): Definition/Des.

<http://freedomdefined.org/Definition/De> [letzter Zugriff: 29.05.2012].

Dewe, Bernd; Weber, Peter (2007): Einführung in moderne Lernformen: Von traditionellen zu computergestützten Lernformen in der europäischen Wissensgesellschaft. Weinheim und Basel: Beltz.

Diepold, Peter (1997): Kinder und Medien: „Gutenberg in Cyberspace“.

http://bildung.freepage.de/cgi-bin/feets/freepage_ext/41030x030A/rewrite/airwolf007/Kapitel1/kapitel1.htm [letzter Zugriff: Abruf: 28.04.2012].

Diedrich, Oliver (2011): Die Woche: Lasst die Daten frei! In: heise online. 11.08.2011.

<http://www.heise.de/open/artikel/Die-Woche-Lasst-die-Daten-frei-1322052.html> [letzter Zugriff: 29.05.2012].

Dietrich, Daniel (2011): Was sind offene Daten?

<http://www.bpb.de/gesellschaft/medien/opendata/64055/was-sind-offene-daten> [letzter Zugriff: 29.05.2012].

dpa (2012): Angebote für Medien: dpa-DataReporting – Der Blick hinter die Zahlen.

<http://www.dpa.com/dpa-DataReporting.197.0.html> [letzter Zugriff: 19.06.2012].

The Election Will Be Tweeted (and Retweeted) (2010). In: The New York Times. 05.06.2010.

<http://www.nytimes.com/interactive/us/politics/2010-twitter-candidates.html> [letzter Zugriff: 15.06.2012].

Elmer, Christina (2012): Data Driven Journalism: Neue Erzählformen mit Datenjournalismus.

<http://www.bpb.de/system/files/pdf/9NGOGE.pdf> [letzter Zugriff: 09.05.2012].

European Journalism Centre (EJC) (2010): Data-driven journalism: What is there to learn? A paper on the data-driven journalism roundtable held in Amsterdam on 24 August 2010.

<http://bit.ly/KCOl4p> [letzter Zugriff: 20.05.2012].

- Focke, Jonathan (2011): Die Nachricht aus der Datenbank.
<http://www.derblindefleck.de/index.php/2011/07/06/die-nachricht-aus-der-datenbank/> [letzter Zugriff: 21.05.2012].
- Fomin, Alexander (2010): Semi-automatische Generierung von REST-Services aus Webseiten.
<http://www.se.uni-hannover.de/pub/File/pdfpapers/Fomin2010.pdf> [letzter Zugriff: 01.06.2012].
- Forum Journalismus und Medien Wien (2012): Crowdsourcing – Die Weisheit der Massen nutzen.
<http://www.fjum-wien.at/wp/kurse/crowdsourcing-die-weisheit-der-massen-nutzen/> [letzter Zugriff: 21.05.2012].
- Frankfurt Gestalten (2012): Häufige Fragen.
<http://www.frankfurt-gestalten.de/haeufige-fragen> [letzter Zugriff: 17.06.2012].
- Gehmlich, Pierre (2012): Schnell, Exklusiv, Transparent.
<http://www.onlinejournalismus.de/2012/02/27/schnell-exklusiv-und-transparent/> [letzter Zugriff: 22.06.2012].
- Gordon, Peter (2007): Data Leakage – Threats and Mitigation.
http://www.sans.org/reading_room/whitepapers/awareness/data-leakage-threats-mitigation_1931 [letzter Zugriff: 06.06.2012].
- Grieß, Andreas (2011): Die Datenzwiebel.
<http://www.youdaz.com/2011/06/20/die-datenzwiebel> [letzter Zugriff: 28.05.2012]
- Grimme Online Award (2012): Preisträger 2011.
<http://www.grimme-institut.de/html/index.php?id=1122#c8530> [letzter Zugriff: 10.06.2012].
- Haase, Alexander (2011): Stories in Data. Das Potential von Daten und ihr Einfluss auf den Journalismus. Mittweida: Hochschule Mittweida.
- Haider, Katharina (2011): Let The Data Speak! In: Digital Journal. 25.03.2011.
<http://digitaljournal.zib21.com/let-the-data-speak/304901/> [letzter Zugriff: 20.05.2012].
- Haller, Michael ⁵2000 (1983): Recherchieren. Konstanz: UVK-Medien.

Hartmann, Robin (2011): Computer Assisted Reporting – Vater und Bruder des Datenjournalismus.

<http://www.youdaz.com/2011/05/11/computer-assisted-reporting-vater-und-bruder-des-datenjournalismus/> [letzter Zugriff: 21.05.2012].

Heidmeier, Markus (2012): Datenjournalisten aller Länder, vereinigt euch!

<http://blog.zeit.de/open-data/2012/02/13/datenjournalisten-aller-lander-vereinigt-euch/> [letzter Zugriff: 09.05.2012].

Hilbert, Martin; Lopez, Priscila (2011): The world's technological capacity to store, communicate and compute information (Abstract).

<http://www.sciencemag.org/content/332/6025/60> [letzter Zugriff: 28.04.2012].

Hoely, Daniel (2011): Scraping: Ist das Kopieren von Daten aus dem Netz erlaubt?

<http://blog.kooptech.de/2011/09/scraping-ist-das-kopieren-von-daten-aus-dem-netz-erlaubt/> [letzter Zugriff: 01.05.2012].

Holovaty, Adrian (2006): A fundamental way newspaper sites need to change.

<http://www.holovaty.com/writing/fundamental-change/> [letzter Zugriff: 20.05.2012].

Holzinger, Alexander; Stephan, Heike; Taylor, Bettina; Thar, Björn (2012): Datenjournal: Werkzeugkasten.

<http://www.datenjournal.de/dossier/werkzeugkasten#texttabellen> [letzter Zugriff: 19.06.2012]

Howe, Jeff (2006a): Crowdsourcing A Definition.

http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html [letzter Zugriff: 15.06.2012].

Howe, Jeff (2006b): The Rise of Crowdsourcing. In: Wired Magazine. 14.06.2006.

<http://www.wired.com/wired/archive/14.06/crowds.html> [letzter Zugriff: 18.06.2012].

IBM, Business Analytics & Optimization Insider (2011): Big Data: Datenregen oder Datensegen?

<http://www-01.ibm.com/software/de/insider/ausgaben/2011/2011-3-1a.html?key=Big%20Data> [letzter Zugriff: 04.05.2012].

Internet & Gesellschaft Co:llaboratory (2010): „Offene Staatskunst“: Bessere Politik durch „Open Government“?

<http://bit.ly/L0fdIk> [letzter Zugriff: 29.05.2012].

Investigate your MP's expenses (2009). In: The Guardian.
<http://mps-expenses.guardian.co.uk/> [letzter Zugriff: 18.06.2012].

Jarvis, Jeff (2008): The press becomes the press-sphere.
<http://buzzmachine.com/2008/04/14/the-press-becomes-the-press-sphere/> [letzter Zugriff: 15.06.2012].

Juskalian, Russ (2008): Interview with Clay Shirky, Part I.
http://www.cjr.org/overload/interview_with_clay_shirky_par.php?page=all [letzter Zugriff: 03.05.2012].

Kaute, Sonja (2011): Datenjournalismus & Open Data: Überblick, Beispiele und Quellen für Journalisten.
<http://stift-und-blog.de/2011/11/datenjournalismus-opendata/> [letzter Zugriff: 21.05.2012].

Kaute, Anna-Lena (2012): „Man muss Ängste überwinden und nicht gleich aufgeben.“ Interview mit Mirko Lorenz zum Thema Datenjournalismus.
<http://stift-und-blog.de/2012/04/interview-lorenz-datenjournalismus/> [letzter Zugriff: 19.06.2012].

Kayser-Brill, Nicholas (2011): Mailinglist: [ddj] RES: data-driven-journalism Digest, vol. 7, issue 5.
<http://lists.okfn.org/pipermail/data-driven-journalism/2011-October/000098.html> [letzter Zugriff: 16.06.2012].

Klinger, Corinna (2012): Crowdsourcing: Guttentag: Die Arbeit der Vielen.
<http://www.datenjournal.de/dossier/beispiele> [letzter Zugriff: 18.06.2012].

Konitzer, Michael (2011): Big Data.
<http://konitzer.wordpress.com/tag/information-overflow/> [letzter Zugriff: 02.05.2012].

Köppli, Stefan (2011): Presseinformation: Neue Cisco-Prognose: Weltweiter IP-Datenverkehr steigt bis 2015 um das Vierfache.
http://www.cisco.com/web/CH/de/press/releases/2011/20110601_vni.html [letzter Zugriff: 01.05.2012].

Krampe, Anna-Lena (2012a): Bestandsaufnahme: Datenjournalismus in Deutschland (Teil 1). In: Datenschaetze. 12.04.2012.
<http://datenschaetze.wordpress.com/2012/04/12/bestandsaufnahme-datenjournalismus-in-deutschland-teil-1/> [letzter Zugriff: 20.05.2012].

Krampe, Anna-Lena (2012b): Bestandsaufnahme: Datenjournalismus in Deutschland (Teil 2). In: Datenschaetze. 23.04.2012.
<http://datenschaetze.wordpress.com/2012/04/23/bestandsaufnahme-datenjournalismus-in-deutschland-teil-2/> [letzter Zugriff: 19.06.2012]

Krampe, Anna-Lena (2012c): Vom CAR zum Datenjournalismus. Wandel des Begriffs und neue Wege. In: Datenschaetze. 05.04.2012.
<http://datenschaetze.wordpress.com/2012/04/05/vom-car-zum-datenjournalismus/> [letzter Zugriff: 20.05.2012].

Krampe, Anna-Lena (2012d): Was ist CAR? In: Datenschaetze. 24.03.2012.
<http://datenschaetze.wordpress.com/2012/03/24/was-ist-car/> [letzter Zugriff: 20.05.2012].

Kreutz, Christian (2010): Frankfurt gestalten. Offene Daten für eine partizipative Lokalpolitik. In: Open Data Network. 01.03.2010.
<http://opendata-network.org/2010/03/frankfurt-gestalten-offene-daten-partizipative-lokalpolitik/> [letzter Zugriff: 17.06.2012].

Kunczik, Michael; Zipfel, Astrid (2001): Publizistik. Ein Studienhandbuch. Köln [u.a.]: UTB.

Landrock, Holm; Zilch, Andreas (2012): Big Data: Eine neue Dimension für Informationen – Teil 3: Tempo und Nutzerzahlen.
<http://www.experton-group.de/research/ict-news-dach/news/article/big-data-eine-neue-dimension-fuer-informationen-teil-3-tempo-und-nutzerzahlen.html> [letzter Zugriff: 06.05.2012].

Lang, Christian (2010): Einsatzpotentiale von Crowdsourcing und Open Innovation für das Customer Experience Management. Heilbronn: GRIN Verlag.

Langer, Ulrike (2011a): Daten rücken in den Fokus: Wie Fachjournalisten davon profitieren können.
<http://bit.ly/JHv4Q3> [letzter Zugriff: 02.05.2012].

Langer, Ulrike (2011b): Datenjournalismus. Journalisten Werkstatt. Salzburg-Eugendorf: Medienfachverlag Oberbauer.

Langer, Ulrike (2011c): Ortsdaten als Chance für Lokaljournalismus. In: Der Standard.at. 21.11.2011.
<http://derstandard.at/1319183374006/Datenjournalismus-Ortsdaten-als-Chance-fuer-Lokaljournalismus> [letzter Zugriff: 19.06.2012].

Langer, Ulrike (2012): Journalismus & Partizipation: Hyperlocal: Chance oder Bedrohung?

<http://diskurs.dradio.de/2012/05/25/hyperlocal-plattformen-journalisten-muessen-ihr-selbstverstandnis-andern/> [letzter Zugriff: 17.06.2012].

Leßmöllmann, Annette (2012): Datenjournalismus: Chance für den Journalismus von morgen.

<http://journalistik-journal.lookingintomedia.com/?p=843> [letzter Zugriff: 21.05.2012].

Lindenberg, Friedrich (2012): Getting Data from the Web.

http://datajournalismhandbook.org/1.0/en/getting_data_3.html [letzter Zugriff: Abruf: 01.06.2012].

Lisakowski, Vera (2012): Wozu Datenjournalismus?

<http://blog.grimme-online-award.de/2012/05/wozu-datenjournalismus/> [letzter Zugriff: 18.06.2012].

Lobmeier, Sabine (2012): Cisco-Studie: Mobiler Datenverkehr steigt weltweit von 2011 bis 2016 um das Achtzehnfache.

<http://globalnewsroom.cisco.com/easyir/DE/de/local/press-release/Cisco-Studie-Mobiler-Datenverkehr-steigt-weltweit-von-2011-bis-2016-um-das-856402.html> [letzter Zugriff: 02.05.2012].

Lokalrundfunktage: 2012: Lorenz Matzat, OpenDataCity Berlin, bei der Local Web Conference 2012.

<http://www.youtube.com/watch?v=ZrV7VDNwjgQ> [letzter Zugriff: 17.05.2012].

Lorenz-Meyer, Lorenz (2011): Was ist Datenjournalismus?

<http://www.scarlatti.de/?p=861> [letzter Zugriff: 17.05.2012].

Macromedia Hochschule für Medien und Kommunikation (2012): Datenjournalismus ein wenig mehr journalistische Unabhängigkeit.

<http://journalismus-mhmk.de/medialab/2012/01/datenjournalismus-ein-wenig-mehr-journalistische-unabhangigkeit/> [letzter Zugriff: 14.06.2012].

Mahmood, Edeela (2011): Was ist Datenjournalismus? Beschreibung und Versuch einer Definition.

<http://www.youdaz.com/2011/04/26/was-ist-datenjournalismus-beschreibung-und-versuch-einer-definition/> [letzter Zugriff: 20.05.2012].

Marks, Peter B. (2009): Journalism changed tonight.

<http://blog.marxy.org/2009/06/journalism-changed-tonight.html> [letzter Zugriff: 15.06.2012].

Matzat, Lorenz (2010a): Data Driven Journalism: Versuch einer Definition.
<http://opendata-network.org/2010/04/data-driven-journalism-versuch-einer-definition/> [letzter Zugriff: 07.05.2012].

Matzat, Lorenz (2010b): Datenjournalismus vor dem Internet: Wetterbericht, Finanzdaten und Co.
<http://www.datenjournalist.de/datenjournalismus-vor-dem-internet-wetterbericht-finanzdaten-und-co/> [letzter Zugriff: 21.05.2012].

Matzat, Lorenz (2010c): Definitionen: OpenData, OpenGovernment, Gov2.0 und Co.
<http://www.datenjournalist.de/definitionen-opendata-opengovernment-gov2-0-und-co/> [letzter Zugriff: 21.05.2012].

Matzat, Lorenz (2010d): Einführung in Datenjournalismus / Data Driven Journalism.
http://www.slideshare.net/wir_sie/einfhrung-in-datenjournalismus-data-driven-journalism [letzter Zugriff: 09.05.2012].

Matzat, Lorenz (2010e): Hyperlocal: Lokaljournalismus und OpenData in den USA und Deutschland.
<http://www.datenjournalist.de/hyperlocal-microlocal-usw-lokaljournalismus-und-opendata-in-den-usa-und-deutschland/> [letzter Zugriff: Abruf: 17.06.2012].

Matzat, Lorenz (2011a): Datenjournalismus: Neue Erzählformen.
<http://www.bpb.de/gesellschaft/medien/opendata/64069/datenjournalismus?p=2> [letzter Zugriff: 10.06.2012].

Matzat, Lorenz (2011b): Daten sind der Treibstoff des Journalismus.
<http://blog.zeit.de/open-data/2011/05/04/daten-journalismus-texastribune/> [letzter Zugriff: 04.05.2012].

Matzat Lorenz (2011c): Dossier Open Data: Datenjournalismus.
<http://www.bpb.de/gesellschaft/medien/opendata/64069/datenjournalismus> [letzter Zugriff: 07.05.2012].

Matzat, Lorenz 2011d: Hyperlokal: Lokale Datenprojekte.
<http://www.bpb.de/gesellschaft/medien/opendata/64073/lokale-datenprojekte> [letzter Zugriff: 17.06.2012].

Matzat, Lorenz (2011e): Data-Blog. Ein Glossar rund um Open Data. In: Zeit online. 13.05.2011.
<http://blog.zeit.de/open-data/2011/05/13/begriff-definition-opengov/> [letzter Zugriff: 17.06.2012].

Matzat, Lorenz (2012a): Datenjournalismus im Lokalen: Alles auf eine Karte.
<http://www.journalist.de/ratgeber/handwerk-beruf/redaktionswerkstatt/daten-journalismus-im-lokalen-alles-auf-eine-karte.html> [letzter Zugriff: 19.06.2012].

Matzat, Lorenz (2012b): Mehr als 1000 Worte.
http://www.screen.tv/pdf/ausgabe_09/02_TREND.pdf [letzter Zugriff: 15.06.2012].

McGregor, Glen (2012): Web scraping: How journalists get their own data.
<http://j-source.ca/article/web-scraping-how-journalists-get-their-own-data> [letzter Zugriff: 01.06.2012].

Medium Magazin (2011): „Man macht sich überprüfbar“.
<http://www.mediummagazin.de/archiv/2011-2/ausgabe-01022011/man-macht-sich-ueberpruefbar> [letzter Zugriff: 09.05.2012].

Meier, Klaus (2011): Journalismus 3.0: Trend Datenjournalismus.
<http://journalistiklehrbuch.wordpress.com/2011/09/09/journalismus-3-0-trend-datenjournalismus/> [letzter Zugriff: 17.05.2012].

Meyer, Philip (2012): The New Precision Journalism.
<http://www.unc.edu/~pmeyer/book/Chapter1.htm> [letzter Zugriff: 17.05.2012].

Neuroth, Oliver (2011): Internetprojekt GuttenPlag: „Schwarmintelligenz“ im Kampf gegen Plagiate. In: tagesschau.de. 19.02.2011.
<http://www.tagesschau.de/inland/guttenplag100.html> [letzter Zugriff: 18.06.2012].

OpenDataCity (2012): Wir sind zweimal für den Grimme Online Award 2012 nominiert.
<http://www.opendatacity.de/wir-sind-zweimal-fur-den-grimme-online-award-2012-nominiert/> [letzter Zugriff: 20.05.2012].

Open Data Network (2010): Datenjournalismus in der Praxis.
http://www.slideshare.net/opendata/data-driven-journalism-in-der-praxis?src=related_normal&rel=7179916 [letzter Zugriff: 20.05.2012].

Open Knowledge Foundation (2012): Was ist Open Data?
<http://opendatahandbook.org/de/what-is-open-data/index.html> [letzter Zugriff: 29.05.2012].

Oswald, Bernd (2011): Die Intelligenz des Schwarms. In: medium online 3 (2011).
<http://www.mediummagazin.de/archiv/2011-2/ausgabe-32011/die-intelligenz-des-schwarms/> [letzter Zugriff: 18.06.2012].

Paderta Damian (2012): OpenData – raumbezogene Daten.
<http://paderta.com/stuff/opendata-raumbezogene-daten-full/> [letzter Zugriff: 29.05.2012].

Palmer, Michael (2006): Data is the New Oil.
http://ana.blogs.com/maestros/2006/11/data_is_the_new.html [letzter Zugriff: 04.05.2012].

Patzwald, Marc (2012): Erzählen nach Zahlen: Datenjournalismus zwischen neuen Formen und alten Ideen. In: epd medien.
<http://www.epd.de/fachdienst/fachdienst-medien/schwerpunktartikel/erz%C3%A4hlen-nach-zahlen> [letzter Zugriff: 18.06.2012].

Pelzer, Claudia (2011): Crowdsourcing Terminologie.
<http://www.crowdsourcingblog.de/blog/2011/02/12/terminologie/> [letzter Zugriff: 21.05.2012].

Plöchinger, Stefan (2012a): Decoder für den Bahn-Stau. Geschichte des Zugmonitors. In: süddeutsche.de. 09.03.2012.
<http://www.sueddeutsche.de/reise/geschichte-des-zugmonitors-decoder-fuer-den-bahn-stau-1.1304468> [letzter Zugriff: 15.06.2012].

Plöchinger Stefan (2012b): Wir speichern, Sie stöbern. In: süddeutsche.de. 10.03.2012.
<http://www.sueddeutsche.de/kolumne/datenjournalismus-projekt-zugmonitor-wir-speichern-sie-stoebern-1.1304376> [letzter Zugriff: 17.05.2012].

Pohlmann, Sonja (2011): Schwarm und Schwärmer. Plagiatsdebatte. In: Der Tagesspiegel. 22.02.2011.
<http://www.tagesspiegel.de/medien/plagiatsdebatte-schwarm-und-schwaermer/3872214.html> [letzter Zugriff: 18.06.2012].

Posch, Eva (2009): Historische Darstellungsformen des amerikanischen Printjournalismus und deren Adaption in Europa anhand des Beispiels Österreich.
http://othes.univie.ac.at/7794/1/2009-11-25_0301721.pdf [letzter Zugriff: 17.05.2012].

Rau, Johannes (2002): Sprache im Zeitalter von Informations- und Kommunikationstechnik.
<http://www.mediendaten.de/fileadmin/Texte/Rau02.pdf> [letzter Zugriff: 28.04.2012].

Reiter, Werner (2012): Was ist das journalistische am Datenjournalismus?
<http://www.twentytwenty.at/2012/03/werner-reiter-was-ist-das-journalistische-am-datenjournalismus/> [letzter Zugriff: 06.06.2012].

Rinne, Solveik; Rennhak, Carsten (2006): Information Overload – warum wir in der Kommunikation neue Wege gehen müssen.
http://www.munich-business-school.de/fileadmin/mbs/documents/working_papers/MBS-WP-2006-05.pdf [letzter Zugriff: 02.05.2012].

Rogers, Simon (2010): Information is Power. In: The Guardian. 24.05.2010.
<http://www.guardian.co.uk/media/2010/may/24/data-journalism> [letzter Zugriff: 29.05.2012].

Rogers, Simon (2011): Wikileaks und der investigative Datenjournalismus. Wie wir beim *Guardian* mit den Wikileaks-Dateien umgehen. In: Geiselberger, Heinrich (Hg.): Wikileaks und die Folgen. Netz – Medien – Politik. Berlin: Suhrkamp.

Rogers, Simon (2012): Change – Wie Daten den Journalismus verändern.
<http://diskurs.dradio.de/2012/02/22/change-wie-daten-den-journalismus-verandern/> [letzter Zugriff: Abruf: 20.05.2012].

Rosenbaum, Martin (2012): Data Stories. In: Data Journalism Handbook 1.0. 2012.
http://datajournalismhandbook.org/1.0/en/understanding_data_5.html [letzter Zugriff: 14.06.2012]

Ruppert, Max (2012): Plag-Wikis: Schnipseljagd. In: Journalist online. 13.02.2012.
<http://www.journalist.de/ratgeber/handwerk-beruf/redaktionswerkstatt/crowd-sourcing-wie-journalisten-plagwikis-nutzen.html> [letzter Zugriff: 18.06.2012].

Scheer, August Wilhelm (2011): Information Overload? Wie die Deutschen mit Nachrichtenflut und Medienvielfalt umgehen. BITKOM 2011.
http://www.bitkom.org/files/documents/BITKOM_Praesentation_Info-Management_31_03_2011%281%29.pdf [letzter Zugriff: 02.05.2012].

Schmidt, Dominik 2011: Prozessjournalismus: Die Mentalität des Unfertigen.
<http://do-s.de/prozessjournalismus-die-mentalitaet-des-unfertigen/> [letzter Zugriff: 15.06.2012].

Schneider, Wolf; Raue, Paul-Josef ⁵2003 (1996): Das neue Handbuch des Journalismus. Reinbek bei Hamburg: Rowohlt.

Schuler, Thomas (2009): Weisse Bescheid. „Crowdsourcing“ im Journalismus. In: sueddeutsche.de. 15.04.2009.
<http://www.sueddeutsche.de/kultur/crowdsourcing-im-journalismus-weisse-bescheid-1.395473> [letzter Zugriff: 18.06.2012].

Schulzki-Haddouti, Christiane (2010a): Titelthema: Datenjournalismus: Spannende Recherche im Netz.
<http://mmm.verdi.de/archiv/2010/03/titelthema-datenjournalismus/spannende-recherche-im-netz> [letzter Zugriff: 29.05.2012].

Schulzki-Haddouti, Christiane (2010b): Analysieren Sie selbst! In: Zeit online. 16.04.2010.
<http://www.zeit.de/politik/deutschland/2010-04/parteispenden-analyse> [letzter Zugriff: 16.06.2012].

Schulzki-Haddouti, Christiane (2010c): Open-Data-Serie über Parteispenden bei ZEIT online.
<http://blog.kooptech.de/2010/04/open-data-serie-ueber-parteispenden-bei-zeit-online/> [letzter Zugriff: 16.04.2010].

Schulzki-Haddouti, Christiane (2011): Dossier Open Data: Die globale Bewegung für offene Daten.
<http://www.bpb.de/gesellschaft/medien/opendata/64063/globale-entwicklung> [letzter Zugriff: 29.05.2012].

Schulzki-Haddouti, Christiane (2012): Open Data: Datenjournalismus: Wenn Zahlen Geschichten erzählen.
<http://blog.zdf.de/hyperland/2012/05/datenjournalismus-wenn-zahlen-geschichten-erzaehlen/> [letzter Zugriff: 16.06.2012].

Seiermann, Stefan (2012): Vergleich von Methoden zur Hypertext-Klassifikation.
www.ke.tu-darmstadt.de/bibtex/attachments/single/112 [letzter Zugriff: 02.05.2012].

Sharky, Clay (2008): Web 2.0 Expo NY: It's Not Information Overload. It's Filter Failure.
<http://www.youtube.com/watch?v=LabqeJEOQyI> [letzter Zugriff: 03.05.2012].

Stanoevska-Slabeva, Katarina (2011): Enabled Innovation: Instruments and Methods of Internet-based Collaborative Innovation.
<http://bit.ly/MZFhmA> [letzter Zugriff: 18.06.2012].

Stevens, Jane (2011): What is a multimedia story?
<http://multimedia.journalism.berkeley.edu/tutorials/starttofinish/choose/>
[letzter Zugriff: 17.05.2012].

Stöcker, Christian (2011): Affäre Guttenberg Netz besiegt Minister. In: Spiegel online. 01.03.2011.
<http://www.spiegel.de/netzwelt/netzpolitik/affaere-guttenberg-netz-besiegt-minister-a-748358.html> [letzter Zugriff: 18.06.2012].

Suppan, Angela (2012): Interactive Story Telling.
<http://www.grin.com/de/e-book/42275/interactive-story-telling> [letzter Zugriff: 10.06.2012].

taz. die tageszeitung (2012): Data-Blog.
<http://blogs.taz.de/open-data/> [letzter Zugriff: 19.06.2012].

Thurner, Thomas (2010): Daten Journalismus – der Weg aus der Journalismuskrise
<http://www.zukunftswelt.at/daten-journalismus-der-weg-aus-der-journalismuskrise> [letzter Zugriff: 22.06.2012].

Torial.com (2012): Datenjournalismus für Anfänger.
<http://www.torial.com/blog/2012/06/datenjournalismus-fur-anfanger/> [letzter Zugriff: 19.06.2012].

Trautmann, Svenja (2011): Scraping – Eine neue Methode im Datenjournalismus.
<http://blog.kooptech.de/2011/08/scraping-eine-neue-methode-im-datenjournalismus/> [letzter Zugriff: 01.06.2012].

Verräterisches Handy. In: Zeit online. 10.02.2011.
<http://www.zeit.de/datenschutz/malte-spitz-vorratsdaten> [letzter Zugriff: 10.06.2012].

Vorratsdatenspeicherung (2012): In: Focus Online.
http://www.focus.de/politik/deutschland/antiterrorkampf/vorratsdatenspeicherung_aid_67215.html 10.06.2012].

Walther, Claudia (2001): Was leisten Infografiken in der Tageszeitung? Untersuchung der Zeitungen „Die Welt“ und „Freies Wort“. Leipzig: GRIN Verlag.

Wikileaks (2012): About: What is Wikileaks?
<http://www.wikileaks.org/About.html> [letzter Zugriff: 06.06.2012].

Wu, Suzanne (2011): How much information is there in the world?
<http://phys.org/news/2011-02-world-scientists-total-technological-capacity.html>
[letzter Zugriff: 28.04.2012].

ZAPP (2011): Recherche: Datenjournalismus.
<http://www.youtube.com/watch?v=poUa8Wu9eTk> [letzter Zugriff: 07.05.2012].

Zeppelin Universität (2010): Open Government Data: Frei verfügbare Daten des öffentlichen Sektors.
<http://bit.ly/JL4C8l> [letzter Zugriff: 29.05.2012].